

Humboldt-Universität zu Berlin

Magisterarbeit

**Möglichkeiten der Strukturmodellierung:
Eine exemplarische Zusammenführung
funktionaler Anforderungen an die
Bereitstellung digitaler Forschungsdaten
für ausgewählte
geisteswissenschaftliche Disziplinen**

Arbeit zur Erlangung des akademischen Grades einer Magistra Artium

Philosophische Fakultät I

Institut für Bibliotheks- und Informationswissenschaft

Vorgelegt von: Frau Maxi Kindling

Matrikel-Nr. 185793

Maxi.Kindling@ibi.hu-berlin.de

Gutachter: 1. Prof. Dr. Peter Schirmbacher
 2. Prof. Dr. Stefan Gradmann

Datum der Einreichung: 3. August 2009

Inhalt

Eidesstattliche Erklärung

1	Einleitung	1
1.1	Thema der Arbeit	1
1.2	Hintergrund	3
1.3	Motivation und Zielstellung	6
1.4	Fokus und Abgrenzung	8
1.5	Vorgehen	9
1.6	Verwendete Literatur.....	11
2	Begriffsbestimmung und Einordnung in den aktuellen Forschungsstand	13
2.1	Begriffsbestimmung „Digitale Forschungsdaten“	13
2.2	eResearch und Virtuelle Forschungsumgebungen	14
2.3	Cyberscholarship	18
2.4	Digitale Geisteswissenschaften	21
2.5	Gegenstand der digitalen Geisteswissenschaften	24
2.6	Zusammenfassung	26
3	Forschungsdaten aus Sicht der Geisteswissenschaften.....	27
3.1	Begriffsbestimmung der Geisteswissenschaften	27
3.2	Die hermeneutische Methode als eine Form des Zugangs zum Gegenstandsbereich der Geisteswissenschaften.....	29
3.3	Die Verschränkung von Instrument und Gegenstand auf semiotischer und sprachlicher Ebene	31
3.4	Digitale Forschungsdaten in den Geisteswissenschaften	33
3.5	Zusammenfassung	36
4	Digitale Forschungsdaten in der wissenschaftlichen Wertschöpfung	38
4.1	Die wissenschaftlichen Wertschöpfung in digitalen vernetzten Umgebungen	38
4.2	Aggregationsstufen digitaler Forschungsdaten	43
4.3	Data Curation	46
5	Betrachtungsebenen	49
5.1	Zeichen, Daten, Information, Wissen.....	49

5.2 Perspektivenwechsel: Beschreibung digitaler Objekte	54
6 Komplexität und Kontext digitaler Objekte	55
6.1 Exkurs: Informationsobjekte im WWW – Hypertext und Multimedia...	56
6.2 Betrachtungsebenen eines digitalen Objekts.....	58
6.3 Beschreibungsebenen digitaler Objekte im OAIS-Modell.....	59
6.4 Komplexe digitale Objekte.....	60
6.5 Strukturelle Merkmale und Beschreibung im OAIS-Modell	62
6.6 Zusammenfassung.....	69
7 Funktionale Anforderungen an die Strukturmodellierung.....	70
7.1 Diskursivität und Bezugnahme	71
7.2 Prinzipien im Umgang mit Forschungsdaten	72
7.3 Referring: Verweise	73
7.3.1 Comparison: Vergleiche.....	74
7.3.2 Provenienz	74
7.3.3 Annotation, Edition, Transkription.....	75
7.4 Persistente Identifizierung.....	76
7.5 Discovery	77
7.6 Interoperabilität	78
8 Möglichkeiten der Strukturmodellierung	81
8.1 Modellverständnis „Strukturmodell“	81
8.2 Modelle in der Informatik	83
8.3 Relationen.....	85
8.4 Modellierung im Bereich der digitalen Geisteswissenschaften	87
8.5 OAI „Object Reuse and Exchange“	89
8.6 Zusammenfassung.....	94
9 Abschlussbetrachtung: Potentiale.....	97
10 Literaturverzeichnis	101
10.1 Ausgewertete Literatur- und Internetquellen.....	101
10.2 Weitere Internetquellen	114
11 Anhang	116
11.1 Abbildungsverzeichnis	118
11.2 Tabellenverzeichnis.....	118

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit zum Thema „Möglichkeiten der Strukturmodellierung: Eine exemplarische Zusammenführung funktionaler Anforderungen an die Bereitstellung digitaler Forschungsdaten für ausgewählte geisteswissenschaftliche Disziplinen“ selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle wörtlichen oder sinngemäßen Übernahmen aus anderen Quellen sind als solche gekennzeichnet.

Berlin, den 3. August 2009

1 Einleitung

1.1 Thema der Arbeit

Im World Wide Web (WWW) findet sich eine Vielzahl von Anwendungen, die die Zusammenführung verteilter Daten ermöglichen. So bietet die Referenzierung mithilfe von „Google Maps“¹ oder „Bing Maps“² die Möglichkeit, geografische Daten miteinander in Beziehung zu setzen. Viele Internetnutzer verknüpfen auf zentralen Plattformen wie Online-Communities diverse dezentral vorhandene Inhalte wie Blogpostings, Fotografien oder Musik. Diese Anwendungen sind stark frequentiert und erfreuen sich großer Beliebtheit. Die Grundidee der Verknüpfung von Inhalten besteht zweifellos ebenso in Hinblick auf digitale Forschungsdaten in virtuellen wissenschaftlichen Arbeitsumgebungen.

Die Benutzungsmöglichkeiten digitaler Forschungsdaten in der Wissenschaftspraxis wurden zu Beginn dieses Jahres in einer gemeinsamen Veranstaltung der Deutschen Initiative für Netzwerkinformation e. V. (DINI) und der Deutschen Forschungsgemeinschaft (DFG) zum Thema „Virtuelle Forschungsumgebungen“ diskutiert.³ Ein Ergebnis der Diskussion über den Erfolg virtueller Forschungsumgebungen war die Aussage, dass dieser im Wesentlichen von ihrer Nutzung abhängt. Die Nutzung der bereit gestellten Ressourcen steht wiederum in Abhängigkeit zu einer forschungsnahen Ausrichtung der Angebote virtueller Forschungsumgebungen sowie der Berücksichtigung fachspezifischer Methoden und Forschungsgegenstände.

¹ Vgl. <http://maps.google.de/>

² Vgl. <http://www.bing.com/maps/>

³ Die Veranstaltung fand unter dem Titel „Förderung der wissenschaftlichen Informationslandschaft in Deutschland – Der Aufbau Virtueller Forschungsumgebungen“ statt. – Vgl. <http://www.dini.de/veranstaltungen/workshops/forschungsumgebungen-2009>

Vor diesem Hintergrund scheint eine intensivere Auseinandersetzung mit disziplinären Anforderungen an die Bereitstellung und kollaborative Nutzung digitaler Forschungsdaten von großer Wichtigkeit. Inwieweit Webangebote wie die Online-Communities als Vorbilder für die Entwicklung virtueller Forschungsumgebungen dienen können, ist nicht Gegenstand der vorliegenden Magisterarbeit. Jedoch wird vor dem Hintergrund der wissenschaftlichen Zusammenarbeit der Ansatz aufgegriffen, Forschungsdaten in digital vernetzten Umgebungen miteinander zu verknüpfen. Er wird unter Berücksichtigung spezifischer Anforderungen der geisteswissenschaftlichen Disziplinen behandelt. Dies ist insbesondere von Interesse, da im Bereich der geisteswissenschaftlichen Disziplinen aktuell über die Entwicklung virtueller Forschungsumgebungen diskutiert und der Wunsch nach wissenschaftlicher Kollaboration und interdisziplinären Projekten geäußert wird.⁴

In dieser Magisterarbeit werden digitale Forschungsdaten als digitale Objekte in vernetzten Umgebungen betrachtet. Eine wesentliche Aussage dieser Arbeit lautet, dass Objekte im digitalen Raum nicht als abgeschlossene Objekte betrachtet werden können, sondern Aggregationen aus Datenobjekten sind. Durch Verknüpfungen wie Verweise und die Bezugnahme von Forschungsdaten aufeinander entstehen komplexe Relationen zwischen den Datenobjekten. Die wohl meistgenutzte Form des Verweises in diesem Zusammenhang ist die Zitation. Relationen müssen maschinenlesbar in einer formalen Sprache ausgedrückt werden, um die korrekte Darstellung der Verweise im digitalen Raum zu ermöglichen, wie nachfolgendes Zitat deutlich macht:

⁴ Vgl. Göttinger e-Humanities-Abschluss-Workshop (DFG) – <http://www.textgrid.de/konferenzen/e-humanities-abschluss-workshop-dfg.html>

„The full potential of compound objects cannot be realized unless the structural and component information is both human understandable and machine interpretable.“⁵

Um wiederum den Austausch und die Zusammenführung verteilter Datenbeständen zu ermöglichen, müssen übergreifende Modelle herangezogen werden, die die Relationen zwischen den Datenobjekten berücksichtigen. Für die konkrete Modellierung der Relationen müssen die Wissenschaftsdisziplinen angepasste Lösungen finden. Diese Masterarbeit liefert eine Problematisierung der fachspezifischen Anforderungen an die strukturelle Modellierung digitaler Forschungsdaten.

1.2 Hintergrund

Das WWW als Distributionsort macht für digitale vorhandene Daten möglich, was hinsichtlich analoger Forschungsmaterialien undenkbar war und wie folgt zugespitzt formuliert wird: Analoge Forschungsmaterialien wie Manuskriptsammlungen oder alte Drucke lagerten in Archiven, Bibliotheken oder Museen lokal verteilt und waren durch unzureichende Erschließung meist nicht nachgewiesen und damit nicht auffindbar. Mit der zunehmenden Digitalisierung von Printmaterialien und kulturellen Artefakten werden sie erschlossen und in digitalen Bibliotheken, Repositorien, Virtuellen Forschungsumgebungen und durch weitere Anbieter digitaler Materialien nachgewiesen. Sie sind über das WWW verfügbar und vor allem nachnutzbar. Über digital vernetzte Umgebungen entstehen neue Zugangsformen zu digitalen Repräsentationen analoger Materialien sowie zu genuin digitalen Materialien für Wissenschaft und Allgemeinheit.

⁵ Cheung et al. (2008) SCOPE: A Scientific Compound Object Publishing..., S. 6

Die intensive Auseinandersetzung mit digitalen Forschungsdaten – im deutschsprachigen Raum wird hier häufig der Terminus „Primärdaten“ verwendet – erfolgt international in jüngster Zeit vor allem unter dem Aspekt der Veröffentlichung von Forschungsdaten und des freien Zugangs zu wissenschaftlichen Ergebnissen („Open Data“) als Pendant zu Open Access im Bereich des wissenschaftlichen Publizierens.

Gründe für diese Bemühungen sind knapper werdende finanzielle Ressourcen der öffentlichen Haushalte, die die Forschung finanzieren und die Vermeidung von redundanter Forschungsarbeit erforderlich machen. Durch eine zunehmende Fragmentierung der Wissenschaft und der damit verbunden Verteilung von technischen und finanziellen Ressourcen und der Forschungsdaten wächst der Wunsch nach der Zusammenführung wissenschaftlicher Disziplinen im Rahmen gemeinsamer Forschungsvorhaben. Kollaboration, interdisziplinäre Forschung und der Austausch von Forschungsdaten finden auf Basis vernetzter Strukturen eine ideale Voraussetzung. Die Vorteilen des sogenannten „Data Sharing“ stellt das britische „UK Data Archive“ unter den folgenden Punkten heraus:

- „encourages scientific enquiry and debate
- enables scrutiny of research outcomes
- facilitates research beyond the scope of the original research
- leads to new collaborations between data users and data creators
- reduces the cost of duplicating data collection
- provides important resources for education and training
- encourages the improvement and validation of research methods
- promotes the research that created the data and its outcomes
- can provide a direct credit to the researcher as a research output in its own right“⁶

⁶ UK Data Archive (2008) Managing and Sharing Data, S. 4

Das Potential des Data Sharing kann nur voll ausgeschöpft werden, wenn die Daten entsprechend nachnutzbar sind. Digitale Forschungsdaten sollen möglichst umfassend ausgewertet werden. Ihre Bereitstellung und der Nachweis helfen weiterhin, die redundante Datenerhebung sowie -erhaltung zu vermeiden.

Die erhobenen Daten sind mitunter einmalig (wie beispielsweise Beobachtungen in der Meteorologie bzw. Wetterdaten) und können nicht repliziert werden. Dies trifft nicht nur auf naturwissenschaftliche Experimente, sondern auch in der geisteswissenschaftlichen Forschung zu: Sprache, Literatur und Kultur lassen sich in ihrem jeweiligen historischen Kontext nicht wiederholen.⁷ Die Umstände ihrer Erhebung oder Generierung – also beispielsweise ein Experiment oder eine Digitalisierungsmaßnahme – sollten reproduzierbar bzw. nachvollzogen werden können, um die Nachnutzbarkeit zu gewährleisten und Forschungsergebnisse ggf. im Sinne der sogenannten „guten wissenschaftlichen Praxis“⁸ reproduzierbar zu machen. Die Bereitstellung von Forschungsdaten aller Wissenschaftsdisziplinen wird meist in Hinblick auf den Urheberrechtsschutz und die Nutzungsrechte von Daten problematisch. Die Frage, wem die generierten Daten und die Ergebnisse eines wissenschaftlichen Arbeitsprozesses eigentlich gehören, kann selten eindeutig beantwortet werden.

⁷ Snow et al. (2008) Considering the User Perspective

⁸ Vgl. DFG (1998) Vorschläge zur Sicherung guter wissenschaftlicher Praxis

1.3 Motivation und Zielstellung

Digitale Forschungsdaten verändern den geisteswissenschaftlichen Forschungsprozess, wie in dem nachfolgenden Zitat ausgeführt wird:

„The materials that humanities scholars use in their work are complex, with shifting shades of meaning that are not easily interpreted. [...] Far from attempting to replace the scholar's mind as the processing device, computer delivery of resources can help to support the process. The complexity of visual devices as a way of enshrining memory and communicating knowledge is something that the ancient world understood very well [...] While much of this knowledge has been lost in the textual obsession of print culture, the graphical interface of the computer screen has helped us reconnect to the world of the visual and recognize that we can relearn a long-neglected vocabulary of interpretation. Digital resources can provide us with a new way to see, and thus to perceive the complexities in the process of interpreting humanities materials. A new way of looking at a text can lead to a way of reading it that is unconstrained by the bindings of the printed medium, even if it leads us back to the pages of a printed book.”⁹

Im Paradigma der „digitalen Geisteswissenschaften“ sind vor diesem Hintergrund neue Forschungsansätze und -fragen, die aus Erkenntnisprozessen auf Grundlage digitaler Forschungsdaten resultieren, sicherlich am interessantesten, wenngleich zumindest in deutschsprachiger Literatur kaum benannt. Die Form der Magisterarbeit setzt Grenzen hinsichtlich des Umfangs und das sehr junge Betrachtungsfeld befindet sich permanent in Bewegung und Veränderung. Entsprechend wird ein Grundproblem abstrakt und aus einer gewissen Distanz heraus betrachtet. Mit der Festlegung des Arbeitsschwerpunkts auf den Bereich der strukturellen Modellierung digitaler Objekte wird schließlich ein Teilaspekt herausgegriffen und mögliche Ansätze zukünftiger Forschung aufgezeigt. Aus dem persönlichen Interesse für den Bereich der Geisteswissenschaften wurde die Eingrenzung auf die Hermeneutik als Methode der Zeichenauslegung gewählt, um zugleich das Verhältnis von Daten und Zeichen näher zu untersuchen.

⁹ Warwick (2004) Print Scholarship and Digital Resources, S. 379

Auf der Grunderkenntnis, dass unterschiedliche wissenschaftliche Disziplinen spezifische Anforderungen an die Repräsentation digitaler Forschungsdaten stellen, basiert nachfolgende Arbeitshypothese:

In den hermeneutisch geprägten Disziplinen der Geisteswissenschaft besteht ein genauer zu bestimmendes Verhältnis zwischen dem Zugang zum Forschungsgegenstand und dem Forschungsgegenstand selbst. Dies wird insbesondere in Hinblick auf die digital vernetzten Forschungsdaten deutlich.

Aus dem zu spezifizierenden Umgang mit digitalen Forschungsdaten werden funktionale Anforderungen an die Modellierung der strukturellen Eigenschaften digitaler Forschungsdaten formuliert.

Weiter ausgeführt beinhaltet die These, dass durch den Einsatz struktureller Modellierung von Daten die Bedeutungsfindung und -produktion unterstützt wird, da die Forschungsdaten durch die Verknüpfung mit anderen Forschungsdaten in einen Kontext eingebunden werden oder eingebunden sind. Die Verknüpfungen digitaler Objekte führen zu einer hohen Komplexität. Da digitale Objekte hinsichtlich ihres strukturellen Kontexts bisher unzureichend erschlossen werden, sind beispielsweise Relationen zwischen zusammengehörigen Objekten nicht immer eindeutig abgebildet. Häufig bleibt daher unklar, auf welches Element sich bestimmte Metadaten beziehen – im Falle von Zeitschriften auf die Zeitschrift oder auf die Artikel oder möglicherweise noch kleinere Einheiten wie Absätze. Die Lösung liegt, so die Vermutung, in einer differenzierteren Modellierung der strukturellen Eigenschaften von digitalen Forschungsdaten. Ziel der Arbeit ist, die Prinzipien, denen Modellierung vor diesem Horizont folgen könnte, anhand funktionaler Anforderungen an die Darstellung struktureller Eigenschaften zu konkretisieren.

Die funktionalen Anforderungen werden u. a. aus Prinzipien der hermeneutisch geprägten Wissenschaftskultur entwickelt. Ausschlaggebend ist, wie sich die Wissenschaftler dem Gegenstand nähern und Erkenntnisse gewinnen. In den

hermeneutischen Disziplinen steht der Aspekt der „Bedeutung“ im Vordergrund. Die Methode setzt maßgeblich auf Reflexion und Interpretation. Wie dies geschieht, steht u. a. in Abhängigkeit zur strukturellen Beschaffenheit der Forschungsgegenstände.

1.4 Fokus und Abgrenzung

Das Thema wird in Abgrenzung zur semantischen Modellierung („Semantic Web“) in Hinblick auf formale Sprachen sowie der formalen und inhaltlichen Auszeichnung digitaler Objekte durch Metadaten abgegrenzt. Das Semantic Web hat die Zuweisung von Begriffen zum Ziel. Insofern ist die Bezeichnung „Semantisches Web“ irreführend, da die formalen Sprachen wie Ontologien den Objekten keine Bedeutung zuweisen, sondern Begriffe. Bedeutung hingegen entsteht durch die Interpretation von Zeichen durch einen Akteur wie in der Arbeit zu zeigen sein wird. Das sogenannte „Semantische Web“ hat das Ziel der semantischen Interoperabilität. Um allerdings überhaupt erkennen zu können, in welcher Verbindung digitale Objekte zueinander stehen, müssen auch strukturelle Eigenschaften betrachtet werden. Die Struktur digitaler Objekte kann sich zum einen auf die Syntax digitaler Datenobjekte beziehen, die beispielsweise das Datenformat (wie PDF oder mp3) kodiert. Eine weitere Betrachtungsebene der Struktur wohnt nicht dem Datenobjekte selbst inne, sondern identifiziert Datenobjekte und beschreibt ihre Verbindungen zu anderen Datenobjekten. Dies geschieht aus technischer Sicht mithilfe der Beschreibung durch formale Sprachen wie z. B. Resource Description Framework (RDF). Das Wissen über die Typen von darzustellenden Relationen wird aus einer konzeptuellen Ebene der Betrachtung digitaler Objekte gewonnen, wie in dieser Arbeit ausgeführt wird. Die Bezeichnungen der Relationen selbst sind inhaltlich motiviert, so dass hier der Anschluss an das semantische Web erfolgt. Ziel der strukturellen Modellierung ist die Herstellung struktureller Interoperabilität, die den Austausch und die Verknüpfung digitaler Objekte aus verteilten Beständen ermöglichen.

Die in dieser Arbeit vorgenommenen Zuspitzungen und Eingrenzungen, insbesondere hinsichtlich der Geisteswissenschaften und der Hermeneutik, verstehen sich als Beitrag zu einer fortlaufenden Diskussion. Die Magisterarbeit versucht weniger allgemeingültige Aussagen zu treffen, sondern hat vielmehr die Auseinandersetzung mit einer Grundthese zum Ziel. Annahmen, die anhand einer Theorie (Hermeneutik) zugrunde gelegt werden, werden auf einen anderen Ansatz (Strukturmodellierung) projiziert, um Parallelen aufzuzeigen. In dieser Weise ist etwa die (aus Gründen einer notwendigen Eingrenzung des Themas) vereinfachte und auf Idealtypen zurückgreifende Differenzierung zwischen Geisteswissenschaft bzw. hermeneutischer Methode von den empirischen Ansätzen der Naturwissenschaften zu verstehen. Sie wurde bewusst gewählt, und die Autorin ist sich der Brisanz dieses Ausgangspunktes bewusst. Jedoch lässt sich das Thema ohne diese Abgrenzung nicht bearbeiten.

1.5 Vorgehen

Die in der Arbeit besprochene Thematik ist zweidimensional zu betrachten: Auf der einen Seite sind die Wissenschaftler als Benutzer wissenschaftlicher Daten zu sehen und auf der anderen Seite die Institutionen, die digitale Forschungsdaten und Dienstleistungen bereitstellen. Letztere sollen für ein breites Spektrum an Forschungsdaten wie Rohdaten, hybride Medienobjekte und Arbeitsinstrumente Lösungen hinsichtlich der Sammlung, Erschließung und Bereitstellung in digitalen vernetzten Umgebungen für die Wissenschaft finden. Welche Institutionen hier in Zukunft verantwortlich sind und wie die Kooperation mit der Wissenschaft aussieht, ist eine Frage, die es noch zu beantworten gilt. Fest steht zunächst, dass für einen nachhaltigen Umgang mit digitalen Forschungsdaten die Arbeitskulturen der unterschiedlichen Wissenschaftsdisziplinen einbezogen werden sollten. Dazu zählt auch das Wissen darüber, wie die Forschungsgegenstände in digitalen Umgebungen strukturell beschaffen sind. Die Darstellung digitaler Forschungsdaten und die entsprechenden Zugangsmöglichkeiten sind in Abhängigkeit zu ihrer strukturellen Beschreibung zu sehen.

Nach einer Begriffsbestimmung digitaler Forschungsdaten und der Herstellung des Bezugsrahmens „e-Research“ und „digitale Geisteswissenschaften“ wird in dieser Arbeit das breite Spektrum der digitalen Forschungsdaten im Bereich der Geisteswissenschaften betrachtet. In diesem Zusammenhang werden die verschiedenen Sichtweisen auf Forschungsdaten als Daten im informatorischen Sinne, als Information in der wissenschaftlichen Wertschöpfung und als Bedeutungsträger im semiotischen Sinne dargestellt.

Anschließend wird im fünften Kapitel die Betrachtungsebene der digitalen Objekte für digitale Forschungsdaten eingeführt. Digitale Objekte werden als logische Einheiten betrachtet. Die Betrachtungsebenen digitaler Objekte werden weiterhin genutzt, um Klarheit über den Begriff der „Struktur“ zu erlangen, von dem in dieser Arbeit die Rede ist. Darüber hinaus wird dargestellt, welche strukturellen Eigenschaften digitale komplexe Objekte haben, um herauszuarbeiten, welche Anforderungen an ihre Beschreibung (mithilfe der Modellierung) bestehen.

Im Anschluss daran werden exemplarisch funktionale Anforderungen benannt, die sich auf den Umgang mit digitalen Forschungsdaten aus einer hermeneutischen Perspektive ergeben. Die funktionalen Anforderungen sind eine Synthese der bis zu diesem Kapitel gewonnenen Erkenntnisse über die Komplexität und den Kontext digitaler Objekte sowie grundlegender Prinzipien im Umgang mit Forschungsdaten, die einigen Literaturquellen entnommen sind. Infolge dessen erfolgt eine Zuspitzung auf zwei wesentliche Aspekte der Strukturmodellierung: die Identifizierung digitaler Objekte sowie die Relationen zwischen digitalen Objekten. In der Auseinandersetzung mit der Wissenschaftspraxis wird deutlich, dass die Prinzipien im Umgang mit Forschungsdaten zumeist nicht auf konkreten Festlegungen beruhen, sondern sich implizit in der Wissenschaftspraxis herausbilden und sich auch verändern können. Dennoch lässt sich disziplinär eine gewisse Stabilität feststellen. Die in dieser Arbeit vorgenommene Betrachtung erhebt keinen Anspruch, jede denkbare Facette wahrgenommen zu haben, sondern möchte einige wesentliche Prinzipien und Anforderungen erfassen und

hinsichtlich ihrer Umsetzung in der strukturellen Beschreibung von digitalen Objekten aufgreifen.

Abschließend wird die Strukturmodellierung im Speziellen behandelt. Es wird das Modellverständnis in dieser Arbeit zusammengefasst, sowie die Modellierung aus Sicht der Informatik und der digitalen Geisteswissenschaften vorgestellt. Der Standard OAI „Object Reuse and Exchange“ wird als Beispiel für die Modellierung von Webressourcen angeführt. In einer Quintessenz dieser Arbeit wird der Stellenwert der Modellierung herausgearbeitet.

Die Magisterarbeit verknüpft Wissen aus der Informatik (technische Sicht auf digitale komplexe Objekte und ihre Struktur), Wissen aus der Bibliotheks- und Informationswissenschaft über die Zusammenhänge von Daten, Information und Wissen sowie eine geisteswissenschaftliche Perspektive durch den Untersuchungsgegenstand der digitalen Forschungsdaten für Geisteswissenschaften und das Heranziehen der hermeneutischen Methode. Die Annahme lautet hier, dass sich diese drei Wissenschaftsbereiche in hervorragender Weise ergänzen können, um den Anforderungen der Forschung im digitalen Zeitalter (e-Research) gerecht zu werden und durch Synergien zu bereichern.

1.6 Verwendete Literatur

Für die Auseinandersetzung mit Forschungsdaten, den Geisteswissenschaften und der wissenschaftlichen Wertschöpfung standen eine Reihe besuchter Veranstaltungen und Vorträge im vergangenen Jahr vorrangig zu den Themen e-Research, Virtuelle Forschungsumgebungen und digitale Forschungsdaten im Mittelpunkt. Daneben wurden die Publikationen von Michael Nentwich sowie das Kompendium „Digital Humanities“ von Schreibman et al. herangezogen. Für die Betrachtungsebenen digitaler Objekte und die Thematik der komplexen digitalen Objekte wurde vorwiegend auf OAIS-Materialien sowie Materialien aus dem Bereich der Langzeitarchivierung zurückgegriffen.

Die Prinzipien des Umgangs mit Forschungsdaten sind als Synthese der aus der Arbeit bis dahin gewonnenen Erkenntnisse zu verstehen. Als Anregung dienten darüber hinaus die Publikationen verschiedener Autoren, die sich in ihrer Forschungstätigkeit den digitalen Geisteswissenschaften bzw. allgemein dem Umgang mit Forschungsgegenständen widmen. Dazu gehören die von Willard McCarty beschriebenen „methodological commons“¹⁰, die von Wendy Pradt Lougee¹¹ zusammengestellten wissenschaftlichen Verhaltensweisen, die von Marcus Beiner erstellten Kategorien der geisteswissenschaftlichen Forschung¹², die Erkenntnisse einer Projektstudie von William S. Brockman et al.¹³ sowie den Ergebnisse einer Studie von Kellie Snow et al. zur Benutzung digitaler Information¹⁴. Die in dieser Arbeit schließlich formulierten funktionalen Anforderungen sind hauptsächlich in Anlehnung an die von John Unsworth¹⁵ zusammengestellten „scholarly primitives“¹⁶ entstanden und exemplarisch aufgeführt.

¹⁰ Vgl. McCarty (2005) Humanities Computing

¹¹ Vgl. Pradt Lougee (2007) Promoting Digital Scholarship

¹² Vgl. Beiner (2009) Humanities

¹³ Vgl. Brockman et al. (2001) Scholarly Work in the Humanities...

¹⁴ Vgl. Snow et al. (2008) Considering the User Perspective

¹⁵ John Unsworth ist Direktor des Informatik-Instituts an der Universität von Urbana Champaign (USA) und einer der wesentlichen Akteure im Bereich der digitalen Geisteswissenschaften.

¹⁶ Vgl. Unsworth (2000) Scholarly Primitives

2 Begriffsbestimmung und Einordnung in den aktuellen Forschungsstand

Die Masterarbeit behandelt exemplarische Anforderungen an die Modellierung digitaler Forschungsdaten für die geisteswissenschaftliche Forschung. Die Thematik der digitalen Forschungsdaten ist dafür in den größeren Forschungskomplex des „e-Research“ und seiner disziplinären Ausprägung der „digitalen Geisteswissenschaften“ einzubetten. Gleichzeitig wird der Bezug zu den Möglichkeiten und Herausforderungen digitaler, vernetzter Arbeitsumgebungen hergestellt, auf deren Grundlage die kollaborative Arbeit mit Forschungsdaten stattfindet.

Die in diesem Kapitel eingeführten Begriffsdefinitionen bilden zusammen mit einer Einordnung in den aktuellen Forschungsstand den Bezugsrahmen für die vorliegende Masterarbeit. Der Fokus liegt auf den organisatorischen und infrastrukturellen Voraussetzungen für die kollaborative Arbeit mit Forschungsdaten.

2.1 Begriffsbestimmung „Digitale Forschungsdaten“

Digitale Forschungsdaten sind alle digitalen Daten, die für den wissenschaftlichen Arbeitsprozess relevant sind. Dieser Prozess beinhaltet die Vorgänge von der Datengenerierung in einem naturwissenschaftlichen Experiment oder einer sozialwissenschaftlichen Erhebung bis hin zur Fixierung von Forschungsergebnissen in Form einer Veröffentlichung. Digitale Forschungsdaten treten in Abhängigkeit der wissenschaftlichen Disziplinen in den verschiedensten Medientypen und Formaten auf.

Dies ist eine relativ allgemein gehaltene Definition und die Behauptung, zu digitalen Forschungsdaten zählten auch wissenschaftliche Veröffentlichungen, sicherlich streitbar. Damit ist die Schwierigkeit des Begriffs „digitale Forschungsdaten“ auf den Punkt gebracht. Es herrscht keine Einigkeit über den Begriffsumfang. Digitale Forschungsdaten sind insbesondere vor ihrem

disziplinären Hintergrund schwer zu bestimmen. Hinsichtlich ihrer Archivierung, des Managements von Forschungsdaten sowie ihrer Bereitstellung ist eine genauere Bestimmung dessen, was Forschungsdaten in einzelnen Disziplinen sind, allerdings unabdingbar.

In der einschlägigen deutschsprachigen Literatur sind allgemein gehaltene Definitionen die Regel.¹⁷ Dies ist kein neuartiges Problem, es potenziert sich allerdings mit der Verortung der Forschung im digitalen Raum und unter Anwendung informationstechnologischer Werkzeuge, wie im Verlauf der Arbeit gezeigt wird. Gleichzeitig wird die allgemeine Definition vor dem disziplinären Hintergrund der Geisteswissenschaft und der wissenschaftlichen Wertschöpfung spezifiziert.

2.2 eResearch und Virtuelle Forschungsumgebungen

Bereits 1999 beschrieb Michael Nentwich die Verlagerung einiger wesentlicher wissenschaftlicher Tätigkeiten in den von Computern und Informations- und Kommunikationstechnologien bestimmten digitalen Raum unter dem Terminus „Cyberscience“.¹⁸ Er bezog sich in seiner perspektivischen Darstellung auf den Wissenschaftsbetrieb in organisatorischer Hinsicht sowie auf die drei wissenschaftlichen Haupttätigkeitsfelder der sogenannten wissenschaftlichen Wertschöpfung¹⁹:

¹⁷ Vgl. DINI e. V. (2009) Positionspapier Forschungsdaten oder Nestor [Hrsg.] (2009) Digitale Forschungsdaten bewahren und nutzen u.v.m. Der Autorin ist keine explizite Definition von „Forschungsdaten“ bekannt. Aus den vorliegenden Materialien zum Thema wird ersichtlich, dass Definitionen entweder wie angedeutet allgemein gehalten werden oder allenfalls aus disziplinärer Perspektive erfolgen.

¹⁸ Vgl. Nentwich (1999) Cyberscience, S. 4

¹⁹ Die Tätigkeitsfelder in Kapitel 4.1 näher beleuchtet.

- der Wissensproduktion,
- der Wissenschaftskommunikation
- und der Vermittlung wissenschaftlicher Information.

Mittlerweile finden etwa ein Dutzend weiterer Bezeichnungen für den Wissenschaftsbetrieb in digitalen vernetzten Umgebungen Verwendung – teilweise als Synonyme, obgleich sie sich in ihren begrifflichen Bezugspunkten unterscheiden. Zu ihnen gehören „e-Science“ für „Electronic Science“ oder auch „Enhanced Science“, „e-Research“, „e-Infrastructure“, „Cyberinfrastructure“, „Virtual Research Environment“. Die begriffliche Diskussion bildet nicht den Gegenstand dieser Arbeit, zeigt aber das zugrunde liegende Problem einer eindeutigen Bestimmung des Forschungshintergrundes auf. Im Folgenden sollen die Begriffe e-Science, e-Research, Cyberinfrastructure, e-Infrastructure sowie Virtual Research Environment (Virtuelle Forschungsumgebungen) näher beschrieben werden.

Der Begriff der e-Science bezieht sich im angloamerikanischen Sprachraum auf die Naturwissenschaften. Wesentliches Merkmal der „e-Science“ ist wissenschaftliche Zusammenarbeit auf Grundlage von Netzwerk- oder GRID-Technologien.²⁰ E-Science schafft auf Basis der Netzwerktechnologie die Verbindung von „data resources, computational resources & people.“²¹ Der Begriff wurde durch den britischen Research Council (maßgeblich John Taylor) im Jahr 2002 ins Leben gerufen und wie folgt definiert: „e-Science is about the global collaboration in key areas of science and the next generation of [computing] infrastructure that will enable it.“²²

²⁰ GRID-Technologie meint das verteilte Rechnen und spielt vor allem dann eine Rolle, wenn aufgrund großer Datenmengen rechenintensive Forschungsprozesse unterstützt werden.

²¹ Frei nach Blanke (2009) Perspektiven der Technikentwicklung für Geistes- und Kulturwissenschaften

²² Vgl. ebd.

Bezieht sich der Begriff „e-Science“ bisher nur auf die Naturwissenschaften, so schlägt Gregory Crane die Ausweitung der „e-Science“ auf den Begriff „e-Research“ vor, der zusätzlich die „humanities“ umfasst.²³ Da sich diese Arbeit speziell mit den Geisteswissenschaften (oder auch „humanities“) befasst, wird im weiteren Verlauf dieser Arbeit der von Crane empfohlene Begriff „e-Research“ verwendet.

E-Research ist demnach als Konzept zu verstehen, das den Transfer der Arbeitsprozesse aller Wissenschaftsdisziplinen beschreibt, in denen die Informations- und Kommunikationstechnologien kollaborativ angewendet werden. Dies wird auf Grundlage einer technologisch vernetzten Infrastruktur möglich, welche in den USA als „Cyberinfrastructure“ bezeichnet wird. Die US-amerikanische National Science Foundation (NSF) beschreibt sie in einem Strategiepapier als Infrastruktur, die die Generierung, Verteilung und Bewahrung von Forschungsdaten, Information und Wissen ermöglicht. Cyberinfrastructure setzt sich aus den folgenden Komponenten zusammen: „hardware for computing, data and networks, digitally-enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools.“²⁴

Michael Fraser stellt den Unterschied der teilweise synonym verwendeten Begriffe „Cyberinfrastructure“ und „Virtuelle Forschungsumgebungen“ („Virtual Research Environment“ (VRE)²⁵) heraus. Virtuellen Forschungsumgebungen beziehen sich auf die Wissenschaftspraxis im Ganzen, während „Cyberinfrastructure“ als die technische Infrastruktur verstanden wird:

²³ Vgl. ebd.

²⁴ NSF (2007) Cyberinfrastructure Vision for 21st Century Discovery, S. 5

²⁵ Synonyme für Virtual Research Environments sind unter anderem „collaboratories“ oder „cyberenvironments“.

„For the most part, the terms VRE, cyber- or e-infrastructure are synonymous. If there is a difference between a VRE and cyber/e-infrastructure then it is this: the VRE presents a holistic view of the context in which research takes place whereas e-infrastructure focusses [sic!] on the core, shared services over which the VRE is expected to operate.”²⁶

Fraser bezeichnet Virtuelle Forschungsumgebungen als eine Verschmelzung von technischer Infrastruktur (Cyberinfrastructure, e-Infrastructure) und dem Angebot von Dienstleistungen für die Wissenschaft. Virtuelle Forschungsumgebungen bauen auf dem Konzept der Virtueller Lernumgebungen (z. B. das Lernmanagementsystem Moodle) und der Idee der e-Science auf.²⁷ Virtuelle Forschungsumgebungen erweitern die infrastrukturell ausgerichtete e-Science um eine forschungsorientierte und fachspezifische Grundstruktur für Dienstleistungen, Werkzeuge und Ressourcen. Sie verbinden die Bereitstellung von Daten und die Unterstützung von Arbeitsabläufen der wissenschaftlichen Tätigkeit in Form von „researchers’ workbench’ tools“.²⁸ Sie sind auf das Engste mit der Entwicklung der e-Science verknüpft:

“The VRE helps to broaden the popular definition of e-science from grid-based distributed computing for scientists with huge amounts of data to the development of online tools, content, and middleware within a coherent framework for all disciplines and all types of research.”²⁹

Von ihrem Ansatz her sind Virtuelle Forschungsumgebungen disziplinär ausgerichtet. Gleichzeitig haben sie den Anspruch, interdisziplinär nutzbar zu sein, und sind daher aus technologischer wie inhaltlicher Sicht generisch konzipiert. Virtuelle Forschungsumgebungen sind Teil der Forschungsinfrastrukturen, die fächerübergreifende Dienstleistungen und Basis-

²⁶ Fraser (2005) Virtual Research Environments

²⁷ Vgl. Voss & Procter (2009) Virtual research environments in scholarly work..., S. 176

²⁸ Vgl. JISC / DDC (2008) Infrastructure Planning and Data Curation

²⁹ Fraser (2005) Virtual Research Environments

Infrastrukturen³⁰ beinhalten.³¹ Durch den weltweiten Zugriff und den Austausch von Forschungsdaten kann Forschung zunehmend in lokal unabhängiger Kooperation und kollaborativ stattfinden. Sie ermöglichen die Integration von Wissenschaftlern und Ressourcen verschiedener Disziplinen, Einrichtungen und Länder.³²

Die vorangehenden Ausführungen stellen den Bezugsrahmen der organisatorischen und infrastrukturellen Voraussetzungen für die kollaborative Arbeit mit digitalen Forschungsdaten in dieser Masterarbeit dar. Die Auseinandersetzung mit digitalen Forschungsdaten in Virtuellen Forschungsumgebungen befindet sich soweit in einem Anfangsstadium der fachlichen Diskussion, die Raum für eine Vielzahl von weiteren Fragestellungen lässt.

2.3 Cyberscholarship

Die Nutzung von wissenschaftlichen Publikationen und Forschungsdaten erreicht vor dem oben beschriebenen Hintergrund eine neue Qualität, die unter dem Begriff „Cyberscholarship“ zusammengefasst wird. „Cyberscholarship“ ergänzt die „Cyberinfrastructure“.³³ Mit diesem Ergebnis schlossen NSF und das britische Joint Information Systems Committee (JISC) einen Workshop, der sich dem Management von „scholarly resources“ auf Basis der Cyberinfrastructure widmete.³⁴

³⁰ Vgl. bspw. die Aktivitäten des European Strategy Forum for Research Infrastructures (ESFRI) auf europäischer Ebene. – URL: <http://cordis.europa.eu/esfri/>

³¹ Vgl. Neuroth (2009) eResearch und die Rolle der wissenschaftlichen Bibliotheken sowie Fraser (2005) Virtual Research Environments

³² Uhlir und Schröder (2007) Open Data for a Global Science, S. OD37

³³ Vgl. Voss und Procter (2009) Virtual research environments in scholarly work..., S. 182

³⁴ Vgl. Arms und Larsen (2007) The Future of Scholarly Communication

Cyberscholarship wird durch zwei Ausprägungen beschrieben:

- „New forms of scholarship“: Von der Behauptung, neue Formen der wissenschaftlichen Tätigkeit seien datengeleitet („data driven science“) wird Abstand genommen. Der Begriff der datengeleiteten Forschung bezieht sich in der ursprünglichen Bedeutung auf die Prozessierung und Verarbeitung großer Datenmengen. Dies spielt jedoch in Disziplinen wie den Geisteswissenschaften eine geringere Rolle als in den Naturwissenschaften. In diesem Sinne wird der Ansatz generalisiert: Digitale Information führt zu neuen Ausprägungen der Forschungstätigkeit. Weiterhin basieren neue Formen der Forschung auf den neuen Kommunikationsmöglichkeiten, die das World Wide Web (WWW) bietet wie Blogs, Wikis, und weitere Möglichkeiten der kollaborativen Arbeit.
- „Content as infrastructure“: Die Sammlungen digitaler Information und entsprechende Software werden zur Forschungsinfrastruktur. Diese Betrachtungsweise setzt voraus, dass die digitale Information sowohl in organisatorischer wie auch in technischer Hinsicht zugänglich ist. Daraus ergibt sich das Aufgabenfeld der Data Curation³⁵: „Now is the time to build the infrastructure to ensure that this digital content remains accessible and usable for current and future scholars.“³⁶

Unter Berücksichtigung dieser Aspekte planen NSF und JISC bis zum Jahr 2015 die Umsetzung folgender strategischer Zielstellung:

³⁵ Das Konzept der “Data Curation” wird in Kapitel 4.3 erläutert.

³⁶ Arms und Larsen (2007) The Future of Scholarly Communication, S. 3

„Ensure that all publicly funded research products and primary resources will be readily available, accessible, and usable via common infrastructure and tools through space and time, and across disciplines, stages of research, and modes of human expression.“³⁷

Im europäischen Wissenschaftsraum sorgen das sechste und siebte EU-Forschungsrahmenprogramm für eine entsprechende infrastrukturelle Versorgung und Ausstattung.³⁸ Grid-Computing wird in Deutschland beispielsweise durch die BMBF-geförderte D-Grid-Initiative³⁹ vorangetrieben. Die Allianz der großen deutschen Wissenschaftsorganisationen widmet sich im Rahmen ihrer Schwerpunktinitiative „Digitale Information“⁴⁰ den Themen Forschungsdaten und Virtuelle Forschungsumgebungen in gleichnamigen Arbeitsgruppen. Weiterhin sind die Aktivitäten der AG Elektronisches Publizieren der Deutschen Initiative für Netzwerkinformation e. V. (DINI) hervorzuheben, die sich inhaltlich wie personell teilweise mit den Allianz-Arbeitsgruppen überschneiden.⁴¹ Die Deutsche Forschungsgemeinschaft (DFG) übernimmt in Bezug auf Forschungsdaten die Rolle einer Empfehlungsgeberin für die Wissenschaft, während die Allianz der Wissenschaftsorganisationen politische und strategische Aussagen zum Umgang mit Forschungsdaten macht. Nach aktuellen Aussagen der Vertreter der Allianz-AG Forschungsdaten lässt sich zusammenfassen, dass momentan „noch mehr Fragen als Antworten“ zum Thema Forschungsprimärdaten in Deutschland bestehen.⁴²

³⁷ Ebd., S. 3. Die „Cyberscholarship Roadmap“ findet sich ebenda, S. 11.

³⁸ Vgl. <http://www.forschungsrahmenprogramm.de/>

³⁹ Vgl. <http://www.d-grid.de/>

⁴⁰ Vgl. Allianz-Initiative „Digitale Information“ (2008) Schwerpunktinitiative „Digitale Information“ der Allianz-Partnerorganisationen

⁴¹ Vgl. bspw. DINI e. V. (2009) Positionspapier Forschungsdaten

⁴² Vgl. Pfeiffenberger (2009) Umgang mit Forschungsdaten

2.4 Digitale Geisteswissenschaften

In jüngster Zeit verzeichnet e-Research einen Wandel von der Entwicklung von Basis-Technologien hin zu einer routinierteren Praxis. Der Ansatz spielt auch in den Sozial- und Geisteswissenschaften eine zunehmende Rolle.⁴³ Die e-Research-Aktivitäten von und für die geisteswissenschaftlichen Disziplinen werden unter der Bezeichnung „digitale Geisteswissenschaften“ oder englischsprachig „e-Humanities“ bzw. „Digital Humanities“ zusammengefasst.⁴⁴

Das Forschungsfeld der heutigen Digital Humanities hat seinen Ursprung bereits in den 1950er Jahren unter der Bezeichnung „Humanities Computing“. Darunter wurde zunächst die rechnergestützte Auswertung und Analyse von zumeist schriftsprachlichen Daten verstanden. Sie geht auf Robert Busa zurück, der als Pionier in diesem Bereich und damit gewissermaßen als Vater des Humanities Computing bezeichnet wird. Er befasste sich mit rechnergestützten Verfahren zur Ermittlung von Konkordanzen in literarischen Texten und der Ausgabe entsprechend spezifizierter Indizes.⁴⁵

Im Jahr 2004 erschien ein Kompendium der Digital Humanities, das ein thematisch breites Spektrum an Beiträgen zu diesem Themenfeld versammelt. Das Kompendium versteht sich selbst als Wegbereiter der „Digital Humanities“ und als die offizielle Festschreibung unter dieser Bezeichnung.⁴⁶ Wie schon erwähnt,

⁴³ Vgl. Voss & Procter (2009) Virtual research environments in scholarly work..., S. 175

⁴⁴ Das begriffliche deutschsprachige Pendant „digitale Geisteswissenschaften“ und das synonym verwendete „e-Humanities“ finden auch ablehnende Stimmen. U. a. werden die Bezeichnungen „IT-basierte Geisteswissenschaften“ oder „computergestützte Arbeitsumgebung für die geisteswissenschaftliche Forschung“ vorgezogen. Die Frage nach der Sinnhaftigkeit einer Metapher wie „digitale Geisteswissenschaft“ ist nicht Gegenstand dieser Arbeit.

⁴⁵ Ein historischer Abriss kann in dieser Arbeit nicht erfolgen. Daher sei auf den Aufsatz von Susan Hockey „The History of Humanities Computing“ verwiesen – vgl. Schreibman, Susan; Siemens, Ray; Unsworth John (2004): A companion to digital humanities. Malden, Mass.: Blackwell Publ. (Blackwell companions to literature and culture, 26), S. 3-19.

⁴⁶ Vgl. Schreibman et al. (2004) A companion to digital humanities

wird auch in den digitalen Geisteswissenschaften seit einigen Jahren an der Umsetzung des e-Research-Ansatzes gearbeitet.

Die organisatorische und technische Infrastruktur für den Bereich der Geisteswissenschaften im Bereich e-Research ist im deutschsprachigen Raum ausbaufähig.⁴⁷ Im Vergleich zur Zahl der Projekte in den Natur- und Ingenieurwissenschaften gibt es im Rahmen der D-Grid-Initiative⁴⁸ weit weniger Projekte für die geisteswissenschaftliche Forschung. Dies kann damit erklärt werden, dass die Geisteswissenschaften nicht im selben Maße mit den gleichen Datenmassen operieren und entsprechend Grid-Technologien benutzen.⁴⁹ Die digitalen Geisteswissenschaften sind bislang vor allem im Bereich der Text- und Editionsforschung konkret wahrnehmbar, wobei der internationale Auszeichnungsstandard der Text Encoding Initiative (TEI)⁵⁰ sowie für den deutschsprachigen Raum Projekte wie TextGrid⁵¹ und im europäischen Rahmen HyperNietzsche⁵² oder Discovery⁵³ hervorzuheben sind.

Über den Kontext des TextGrid-Projekts hinaus sollen in Deutschland die Kompetenzen im Bereich digitaler Geisteswissenschaften gebündelt und institutionalisiert werden. In Großbritannien und den USA wurde die Institutionalisierung wesentlich früher umgesetzt. So wurden in Großbritannien bereits seit Mitte der 1990er geisteswissenschaftliche Kompetenzzentren

⁴⁷ DINI e.V. (2009) Positionspapier Forschungsdaten, S. 2

⁴⁸ Vgl. <http://www.d-grid.de/>

⁴⁹ Klump (2009) Digitale Forschungsdaten, S. 109

⁵⁰ Vgl. <http://www.tei-c.org/>

⁵¹ Vgl. <http://www.textgrid.de/>

⁵² Vgl. <http://www.hypernietzsche.org/>

⁵³ Vgl. <http://www.discovery-project.eu/home.html>

aufgebaut. Die Förderung des britischen Arts and Humanities Data Service (AHDS)⁵⁴ endete allerdings im März 2008.⁵⁵

Der Schwerpunkt der aktuellen deutschen Initiativen liegt im internationalen Vergleich – insbesondere mit Großbritannien und Nordamerika – stärker in den Bereichen der organisatorischen und technischen Infrastruktur und der Entwicklung von Dienstleistungen und Werkzeugen als in der Auseinandersetzung mit inhaltlichen Fragen und neuen Forschungsansätzen. Katja Meffert hat in einer State-of-the-Art-Analyse im Rahmen des einjährigen DFG-Projekts „Konzeptionelle Entwicklung einer Forschungsinfrastruktur für die e-Humanities in Deutschland“ eine Übersicht zu nationalen und internationalen e-Humanities-Zentren, Initiativen und Ansätzen erstellt.⁵⁶ Auf Grundlage dieser Studie wurde das Schwerpunktprogramm (SPP 1442) „Digital Humanities. Research methods and practices in the context of new infrastructures“ bei der DFG beantragt. Im Mittelpunkt dieser Aktivitäten stand die Georg-August-Universität Göttingen, die in Zusammenarbeit mehrerer Fakultäten, der Staats- und Universitätsbibliothek Göttingen, der Akademie der Wissenschaften zu Göttingen, der Herzog August Bibliothek Wolfenbüttel und der Max Planck Digital Library ein deutsches „Center for Digital Humanities“ mangels einer deutschen „Adresse für e-Science“ begründen wollte. Dieses sollte sich in die internationale „Alliance of Digital Humanities Organizations“ (ADHO)⁵⁷ eingliedern.⁵⁸ Das Schwerpunktprogramm wurde durch die DFG nicht angenommen.

⁵⁴ Vgl. <http://ahds.ac.uk/>

⁵⁵ Das AHDS wird nun nicht länger als nationale, sondern institutionelle Einrichtung „Center for e-Research (CeRch)“ am Londoner Kings College weiterbetrieben. Vgl. <http://www.kcl.ac.uk/iss/cerch>

⁵⁶ Vgl. Meffert (2009) State-of-the-Art-Analyse; eine angekündigte Zusammenfassung der Studie in der Fachzeitschrift „Bibliothek: Forschung und Praxis“ lag zum Abschluss der Magisterarbeit noch nicht vor.

⁵⁷ Vgl. <http://digitalhumanities.org/>

2.5 Gegenstand der digitalen Geisteswissenschaften

Forschungsgegenstände der digitalen Geisteswissenschaften sind digitale Repräsentationen menschlicher „Hervorbringungen“ wie Sprache oder kultureller Artefakte wie Musik und Kunst. Mit dem Aufkommen des World Wide Web und der damit verbundenen Verbreitung digitaler Repräsentationen haben die digitalen Geisteswissenschaften einen Höhepunkt erreicht.⁵⁹ Die den Geisteswissenschaften zugeordneten Disziplinen sind über den gemeinsamen Forschungsgegenstand verbunden:

„[They] share a focus on the examination of artifactual evidence of that which makes us human, [...] share a number of commonly held assumptions about the way in which such examination is carried out, both with and without the assistance of the computer. Widely spread through the digital humanities community is the notion that there is a clear and direct relationship between the interpretative strategies that humanists employ and the tools that facilitate exploration of original artifacts based on those interpretative strategies; or, more simply put, those working in the digital humanities have long held the view that application is as important as theory. Thus exemplary tasks traditionally associated with humanities computing hold the digital representation of archival materials on a par with analysis or critical inquiry, as well as theories of analysis or critical inquiry originating in the study of those materials.”⁶⁰

Eine explizite Auseinandersetzung mit den digitalen Forschungsgegenständen für Geisteswissenschaftler findet sich im nachfolgenden Kapitel 3 „Forschungsdaten aus Sicht der Geisteswissenschaften“.

⁵⁸ Vgl. Diskussion im Rahmen des e-Humanities-Abschluss-Workshops – URL: <http://www.textgrid.de/konferenzen/e-humanities-abschluss-workshop-dfg.html>

⁵⁹ Vgl. Schreibman et al. (2004) *A companion to digital humanities*, S. xxiii

⁶⁰ Schreibman et al. (2004) *A companion to digital humanities*, S. xxv

Susan Schreibman et al. fassen die Handlungsfelder der digitalen Geisteswissenschaften hinsichtlich des Umgangs mit digitalen Forschungsgegenständen wie folgt zusammen:

- „[...] a shared focus on preserving digital artifacts (written, painted, carved, or otherwise created), that which is left to us by chance (ruins, and other debris of human activity), or that which has been near-impossible to capture in its intended form (music, performance, and event),
- [...] wishing to preserve these artifacts, what we might now call early forms of data management, to re-represent and manipulate them to reveal properties and traits not evident when the artifact was in its native form,
- [...] concerns itself with the creation of new artifacts which are born digital and require rigorous study and understanding in their own right.”⁶¹

Welche Konsequenzen aus den Handlungsfeldern der digitalen Geisteswissenschaften für die Bereitstellung digitaler Forschungsdaten für geisteswissenschaftliche Disziplinen zu ziehen sind, ist Thema der Arbeit.

⁶¹ Schreibman et al. (2004) A companion to digital humanities, S. xxiv

2.6 Zusammenfassung

Die digitalen Geisteswissenschaften werden für diese Arbeit anhand der folgenden Schwerpunkte zusammengefasst, die zum einen in diesem Kapitel dargestellt wurden und in den nachfolgenden Kapiteln ausgeführt werden:

- Der Bezugspunkt e-Research ist heranzuziehen: die Bereitstellung technischer und organisatorischer Infrastrukturen für die kollaborative Arbeit und den Austausch von Forschungsdaten. Die Infrastruktur ist noch nicht auf demselben Stand wie beispielsweise in den Naturwissenschaften.⁶²
- Die Aufbereitung und Darstellung digitaler Forschungsdaten, die in verschiedenen Medientypen, Datenformaten und Komplexitätsstufen vorliegen und darüber hinaus bislang stark am Paradigma der Print-Medien orientiert sind.
- Der wissenschaftliche Erkenntnisprozess und der Umgang mit Forschungsdaten aus einer disziplinären Perspektive sowie unter Beachtung einer methodischen Ausrichtung.

⁶² Crane (2008) Repositories, Cyberinfrastructure and the Humanities

3 Forschungsdaten aus Sicht der Geisteswissenschaften

Die vorliegende Arbeit hat die geisteswissenschaftlichen Disziplinen zum Gegenstand. Die Abgrenzung von den Naturwissenschaften erfolgt hauptsächlich auf Grund der folgenden drei Aspekte:

- dem Gegenstandsbereich der Forschung,
- der Form des Zugangs zum Forschungsgegenstand (Methode) sowie
- des Verhältnisses von Instrument⁶³ und Gegenstand.

Nachfolgend wird der Gegenstandsbereich der Geisteswissenschaften für den Kontext dieser Arbeit festgelegt. Es erfolgt eine Eingrenzung durch die Konzentration auf die hermeneutische Methode als einer Form des Zugangs zu geisteswissenschaftlichen Forschungsgegenständen. Die semiotische und sprachliche Verschränkung von Gegenständen und Instrumenten in der geisteswissenschaftlichen Forschung werden im Anschluss untersucht. Zum Abschluss des Kapitels erfolgt die Verortung digitaler Forschungsdaten in den Geisteswissenschaften.

3.1 Begriffsbestimmung der Geisteswissenschaften

Unter „Geisteswissenschaften“ werden in dieser Arbeit all jene wissenschaftlichen Disziplinen verstanden, die sich analysierend, interpretierend und verstehend mit dem impliziten und expliziten Wissen beschäftigen, das der menschliche Geist hervorbringt: Sprache, Geschichte in Form der Erinnerung, Kultur, Kunst, Musik und Literatur. Die als „Geisteswissenschaften“ zusammengefassten Disziplinen ähneln sich sowohl methodisch als auch in ihrem Gegenstandsbereich. Die Gegenstände sind Objekte einer wissenschaftlichen Untersuchung, wenn sie unter

⁶³ Der Begriff „Instrument“ umfasst die Mittel und Kanäle der Kommunikation, die unter Anwendung einer Methode dem Zugang zum Forschungsgegenstand dienen.

Zuhilfenahme einer Methode oder eines Instruments betrachtet werden.⁶⁴ In den Geisteswissenschaften werden Forschungsgegenstände oft als Quellen, die Geisteswissenschaften mitunter auch als „Quellenwissenschaften“ bezeichnet.

Während im deutschsprachigen Raum der Begriff Geisteswissenschaften⁶⁵ verwendet wird, so entspricht dies im angloamerikanischen Sprachraum den *humanities*, im französischen Sprachgebiet *les sciences humaines* oder *les lettres*. Mit der Bezeichnung *humanities* waren im ursprünglichen Gebrauch jene Fachdisziplinen gemeint, die sich dem antiken Menschenbild widmeten. Dies wurde später in einem weiteren Sinne aufgefasst und auf verschiedene Geistes- und Sozialwissenschaften ausgedehnt.⁶⁶

Der Philosoph Reinhardt Brandt erklärt die Geisteswissenschaften in Abgrenzung zu den Naturwissenschaften. Dabei konzentriert er sich im nachfolgenden Ausschnitt auf den Umgang mit dem Forschungsgegenstand:

„[Geisteswissenschaftler] gehen vergleichend, gegenüberstellend vor, während der Naturwissenschaftler und Mathematiker seinen funktionalen Zusammenhang direkt fokussiert. Der Naturwissenschaftler braucht *einen* Projektor, in den Hörsälen der Kunstgeschichte stehen dagegen *zwei*, weil man Bildwerke in ihrer Eigentümlichkeit dadurch bestimmt, dass man sie mit geeigneten Kontrastwerken vergleicht. [...] Die Naturwissenschaften vergessen ihre Geschichte und brauchen nur die Literatur der zwei bis drei Jahre oder Monate oder Wochen. Wir dagegen begreifen unsere Geschichte genetisch und beginnen unsere Untersuchungen und Vorträge grundsätzlich mit dem ersten Gang der Ilias und noch früher, ab ovo, oder mit dem Buch der Bücher, der Bibel. [...] Die Rede und die Dichtung, die Komposition und das Malwerk werden von einer natürlichen und geschulten Urteilskraft

⁶⁴ Vgl. Struve (2005) Von der Wirklichkeit zur Wissenschaft, S. 12f.

⁶⁵ Im deutschsprachigen Raum wird zunehmend die Bezeichnung „Kulturwissenschaften“ verwendet, um die Überzeugung auszudrücken, „dass in unserer Zeit als das spezifisch Menschliche und in diesem Sinne Humane weniger ein antikes Ideal, sondern ganz allgemein das Hervorbringen kultureller Leistungen anzusehen sei.“ (Vgl. Schütt (2005) Der „Geist“ in den Geisteswissenschaften, S. 70). Die Verwendung von „Kulturwissenschaft“ anstelle von „Geisteswissenschaft“ ist jedoch umstritten.

⁶⁶ Vgl. Schütt (2005) Der „Geist“ in den Geisteswissenschaften, S. 68ff.

bestimmt, die sich nicht durch Zählen, Messen und Wiegen beeindrucken lässt. [...] [Geisteswissenschaftler] erkennen ihren Inhalt in seiner Eigentümlichkeit durch die gelehrte Zuwendung auf das, was er ist, wie er entstand und was er in einer ihn kennzeichnenden, ihn charakterisierenden Weise nicht ist. Das Erkennen ist somit substantiell, nicht funktional, bezogen auf die einmalige Sache selbst, nicht auf repetierbare Relationen. Das Erkennen läuft auf eine *Bewertung* der Relevanz des untersuchten Komplexes innerhalb der zugehörigen Konstellationen und vielleicht ihrer Nachwirkungen hinaus.“⁶⁷

Dieses Zitat macht die Herangehensweise an den Forschungsgegenstand deutlich, die im Kapitel 7 „Funktionale Anforderungen an die Strukturmodellierung“ näher ausgeführt wird.

3.2 Die hermeneutische Methode als eine Form des Zugangs zum Gegenstandsbereich der Geisteswissenschaften

In Anlehnung an eine dualistische Sicht auf die Wissenschaft⁶⁸ wird die Hermeneutik als geisteswissenschaftliche Methode der „Forschungsquellen verstehenden“ Wissenschaften von den Naturwissenschaften abgegrenzt, für die das „Erklären“ charakteristisch ist. Damit erfolgt eine stark vereinfachte Gegenüberstellung von Hermeneutik und Empirie, die allein den Zwecken dieser Arbeit dienlich sein soll.⁶⁹ Die Konzentration auf die hermeneutische Methode ist in Hinblick auf die Formulierung funktionaler Anforderungen an die Modellierung digitaler Forschungsdaten als Maßnahme der Eingrenzung zu verstehen.

⁶⁷ Brandt (2005) Zustand und Zukunft der Geisteswissenschaften, S. 29f.

⁶⁸ Angefangen bei dem deutschen Psychologen und Philosophen Wilhelm Dilthey, der Anfang des 20. Jahrhunderts den Dualismus zwischen Natur- und Geisteswissenschaft prägte und als Begründer der Geisteswissenschaften gilt.

⁶⁹ So lässt sich exemplarisch anführen, dass es auch geisteswissenschaftliche Disziplinen wie die Korpuslinguistik gibt, die sich dem Gegenstand „Sprache“ empirisch nähern.

Der Begriff der Hermeneutik (griech. hermēneuein *auslegen, deuten, kundgeben, erklären*) wird für die Auslegung (Interpretation) von Zeichen verwendet. Dabei kann alles Zeichen sein, was interpretiert werden kann: „[...] Gegenstand der Auslegung können prinzipiell alle Lebensäußerungen sein, wenn auch traditionell sprachliche Bekundungen (und darunter wiederum schriftliche Verlautbarungen) im Vordergrund stehen (Texthermeneutik).“⁷⁰ Im Anschluss an den anfangs festgelegten Gegenstandsbereich der Geisteswissenschaften wird für diese Arbeit ein sehr weit gefasster Begriff von „Zeichen“ verwendet, die interpretiert werden. Dieser beinhaltet u. a. Bilder, Texte, Musik.

Aus Sicht der Semiotik (griech. semeion *Kennzeichen*)⁷¹ haben Zeichen die Eigenschaft, auf etwas anderes als sich selbst zu verweisen. Sie verweisen auf eine Bedeutung. Unter Anwendung der hermeneutischen Methode werden Zeichen interpretiert, um herauszufinden wofür ein Zeichen steht bzw. worauf es verweist: „While research in the purposive sense of the term brings to the scholar’s attention texts and facts, it is interpreting or structuring these pieces that creates meaning.“⁷²

Für eine Interpretation ist stets ein Akteur nötig, der das Zeichen als solches erkennt oder bestimmt. Bis zu diesem Zustand sind sie nur potentielle Zeichen. Sie sind nicht per se Bedeutungsträger, sondern nur mit Hilfe eines Akteurs. Die Pragmatik ist die Dimension des Zeichens, die einen Akteur zum Handeln

⁷⁰ Mittelstraß [Hrsg.] (2005) Enzyklopädie Philosophie und Wissenschaftstheorie, S. 364

⁷¹ Grundlage bildet insbesondere die strukturalistische Semiotik nach Ferdinand de Saussure und das semiotischen Zeichenmodell. Es unterscheidet den Signifikanten (frz. signifiant) als Laut oder Ausdrucksform eines Zeichens (im Fortlauf der Arbeit ‚Zeiger‘) von der Bedeutungsseite des Zeichens (im Fortlauf der Arbeit ‚Gezeigtes‘), dem Signifikat (frz. signifié). Die Verbindung von Laut- und Bedeutungsform ist mit Ausnahme lautmalerischer Bezeichnungen arbiträr.

⁷² Brockman (2001) Scholarly Work in the Humanities, S. 24

(Erkennen/Bestimmen) von Zeichen auffordert. In der Semiotik wird die pragmatische Komponente auch als Handlungszweck bezeichnet.⁷³

In diesem Zusammenhang sei auf den Ansatz des „Semantic Web“ verwiesen, dessen Ziel ist, die Erkennung von Zeichen zu formalisieren, d. h. einen künstlichen Akteur zu schaffen, der erkennt und bestimmt, was Zeichen sind und in welchem Bedeutungskontext sie repräsentiert sind.

3.3 Die Verschränkung von Instrument und Gegenstand auf semiotischer und sprachlicher Ebene

„The process of research in the humanities and the fruits of that research are closely intertwined. In fact, the results of research may be inseparable from the activity of research and the writing of its interpretation [...]“⁷⁴

Diese Verschränkung von Gegenständen und Instrumenten lässt sich nicht nur für den Forschungsprozess, sondern auch für den Gebrauch von Zeichen ausmachen.

Die Trennung der Geistes- und Naturwissenschaften erfolgt im Verlauf der Arbeit anhand der diskursbestimmten Instrumente und Gegenstände in den hermeneutisch geprägten Geisteswissenschaften. Aus semiotischer Sicht weisen sie einen besonderen Gebrauch von Zeichen auf.

Die Naturwissenschaften verfertigen Annahmen über einen Forschungsgegenstand auf der Grundlage nicht-diskursiven, sondern generierten oder erhobenen Datenmaterials. Sie werden daher als datengeleitet bezeichnet. Die hermeneutisch geprägten Wissenschaften sind sowohl in den Aktivitäten der Wissensproduktion, als auch des wissenschaftlichen Kommunikationsprozesses selbstreflexiv. Die wissenschaftliche Kommunikation umfasst das Prozessieren von repräsentiertem

⁷³ Vgl. Zimmermann (2004) Information in der Sprachwissenschaft, S. 707

⁷⁴ Brockman et al. (2001) Scholarly Work in the Humanities

Wissen in Form diskursiven Überprüfens, Verfeinerns und Verknüpfens.⁷⁵ Erkenntnisse aus diesem Prozess werden in einer zeichenbasierten Form dargestellt. Der Prozess der Erkenntnisproduktion wie auch die Erkenntnis selbst können jeweils Instrument und Gegenstand der Reflexion sein. Ausdruck dieser Verzahnung ist ihr diskursiver Charakter, weshalb sie auch als „Diskurswissenschaften“ bezeichnet werden.⁷⁶

Die gegenseitige Bezugnahme von Instrument und Gegenstand wird bei wissenschaftlichen Kommunikationsprozessen in vernetzten Umgebungen in technischer Hinsicht ausgedrückt. Publikationen als Forschungsergebnis stehen so formal durch die Verlinkung („Zeiger“) ihrer Forschungsgegenstände oder auf einzelne Aktivitäten der Wissensproduktion (z. B. eine Annotation) in einer Relation zu ihrem Inhalt („Gezeigtes“). Diese Betrachtung lässt sich auch umkehren. In Anlehnung an Jan Christoph Meister können beispielsweise Fragen, die den Forschungsgegenstand betreffen, nicht allein Instrument sein, sondern selbst zum Gegenstand werden. Wird einem Text eine semiotische Funktion als Träger sprachlicher Bedeutung zugeordnet („Zeiger“), so ist nicht er selbst der Forschungsgegenstand, sondern die auf ihn bezogenen Fragen:

„Fragen also, die die Genese, Authentizität, Bedingungshaftigkeit, Regularität und Funktionalität von Texten als Bedeutungsträger betreffen.“⁷⁷

Weiter führt Meister aus:

„Fragen nun werden *traditionsbildend*, indem sie zur Entwicklung von Methoden stimulieren. So gibt es zum Beispiel in den Textwissenschaften Methoden, mit denen Fragen hinsichtlich der Authentizität von Texten bearbeitet werden können, darunter etwa die der Stilanalyse oder der

⁷⁵ Nentwich (1999) Cyberscience, S. 5f.

⁷⁶ Vgl. Gradmann (2004) Vom Verfertigen der Gedanken, S. 58f.

⁷⁷ Meister (2005) Projekt Computerphilologie, S. 320

Oeuvregeschichte. Andere Methoden richten sich primär auf die Erforschung von Strukturmerkmalen von Subklassen von Texten; als Beispiel sei die Narratologie erwähnt, die sich speziell der Teilmenge der erzählenden Texte widmet. Fraglos die bekannteste und wichtigste philologische Methode aber ist die Hermeneutik, die die geregelte Explikation von Textbedeutungen zum Gegenstand hat.“⁷⁸

Die Verschränkung von Instrument und Gegenstand lässt sich neben der semiotischen Ebene auch hinsichtlich der sprachlichen Verfasstheit von Aussagen und Forschungsgegenständen feststellen.⁷⁹ In den empirisch geprägten Naturwissenschaften werden Sprachaussagen („Zeiger“) etwa über Experimentergebnisse („Gezeigtes“) getroffen. In den hermeneutisch geprägten Geisteswissenschaften können diese Sprachaussagen selbst zum Objekt der Untersuchung (also als Zeichen interpretiert) werden. Das bedeutet, dass keine klare Trennung zwischen Aussagen und Gegenständen vorgenommen werden kann.

3.4 Digitale Forschungsdaten in den Geisteswissenschaften

Die Problematik der Begriffsbestimmung „digitaler Forschungsdaten“ und die damit einhergehenden Komplikationen wurden bereits aufgezeigt und werden im nachfolgenden Kapitel näher erklärt. Als digitale Forschungsdaten für die Geisteswissenschaften werden in dieser Arbeit all jene genuin digitalen Daten sowie die digitale Repräsentationen analoger Quellen bezeichnet, die gemäß der oben formulierten Definition Gegenstände geisteswissenschaftlicher Forschung und damit Diskursinhalte sein können.

⁷⁸ Ebd.

⁷⁹ Dieser Ausdruck stammt von Struve (2005) Von der Wirklichkeit zur Wissenschaft, S. 12. Zusätzlich kann die Frage gestellt werden, inwieweit sprachliche Aussagen eine wissenschaftliche Erkenntnis beschreiben können, d. h. ob das Medium Sprache das richtige Mittel zur Darstellung ist (siehe *linguistic turn*), ebd.

Digitale Forschungsdaten für die geisteswissenschaftlichen Disziplinen sind heterogen und komplex in Bezug auf ihre Erscheinungsformen und Formate.⁸⁰ Die nachfolgenden Aussagen über digitale Forschungsdaten für die Geisteswissenschaften zeigen diesen Umstand deutlich auf:⁸¹

- heterogene Datentypen wie Video, Audio, Text, 3-D-Objekte, Bilder,
- heterogene, z. T. nicht-standardisierte und proprietäre Datenformate⁸²,
- komplex hinsichtlich der dateneigenen Struktur,
- komplex hinsichtlich der Struktur digitaler Sammlungen und Daten-Aggregation,
- komplex hinsichtlich der semantischen Relationen zwischen den Daten,
- meist eingeschränkter Zugriff auf Daten und dazugehörige Metadaten (oftmals auf physischen Datenträgern),
- durch mangelhafte und nicht interoperable Aufbereitung der Daten meist kaum nachgewiesen und auffindbar,
- kaum Strategien für langfristige Verfügbarkeit,
- kaum vorhandene oder nur unzureichende Werkzeuge zur Datenbearbeitung mit Ausnahme der Text- und Editionsforchung,
- größer werdende Korpora an digitalem Sprachmaterial bspw. durch Digitalisierungsprojekte.

⁸⁰ Vgl. u. a. Blanke (2009) Perspektiven der Technikentwicklung für Geistes- und Kulturwissenschaften oder Neuroth (2009) eResearch und die Rolle der wissenschaftlichen Bibliotheken. Mitunter wird ihnen aus der Sicht der naturwissenschaftlichen Fächer gar eine chaotische und unorganisierte Datenlage nachgesagt (vgl. Rugemer (2008) Kinetik der Kulturen, S. 29).

⁸¹ Vgl. u. a. Neuroth (2008) Research Data Curation in eHumanities; Deegan und Tanner (2004) Conversion of Primary Sources, S. 488; Schoepflin, Urs (2009) Open-Access-Forschungsinfrastruktur für Kulturerbe

⁸² Empfehlungen zu Datenformaten zur Langzeitarchivierung von Forschungsdaten gibt es beispielsweise durch das UK Data Archive: UK Data Archive (2008) Managing and Sharing Data

Die genannten Eigenschaften machen deutlich, dass die Auseinandersetzung mit digitalen Forschungsdaten eine technische und strategische Herausforderung für Datenanbieter und Wissenschaftler ist, die eine intensive Beschäftigung mit jedem einzelnen der genannten Aspekte erforderlich macht. Die Herausforderungen werden durch die britische „E-Science in the Arts & Humanities“-Initiative unter der Bezeichnung „the data grand challenge“ zusammengefasst. Daneben beschreibt sie die „computational grand challenge“, die folgende Aspekte umfasst:

- „Retrieval and delivery of information of semi-structured information in heterogeneous formats
- Fuzzy matching
- Modelling of (past) human activity: Computing statistical models from complex natural (language) data
- Automated annotation
- Digital Art“⁸³.

In Ergänzung dieser Herausforderungen bieten digitale vernetzte Umgebungen wie das World Wide Web (WWW) ideale technische Voraussetzungen, um Forschungsdaten mithilfe einer Adressierung beispielsweise durch Hyperlinks zu verknüpfen. So können Forschungsgegenstände zum einen in neue inhaltliche Zusammenhänge gebracht werden. Zum anderen können diese Zusammenhänge anhand des Verfolgens von Hyperlinks „entdeckt“ werden. Dies stellt perspektivisch einen wesentlichen Vorteil von digital vorhandenen Forschungsdaten dar. Während sich das Lesen und Interpretieren von Texten und anderen Untersuchungsgegenständen nicht einfach deckungsgleich auf das Benutzen von digitalen Ressourcen übertragen lässt und Print- bzw.

⁸³ Blanke (2006) E-Science in the Arts+Humanities

Analogmaterialien weiterhin eine Rolle spielen werden⁸⁴, bieten digital vernetzte Umgebungen hier einen klaren Vorteil: Eine Zusammenstellung von Forschungsdaten kann beispielsweise bedeuten, dass ein Datensatz in einem Repositorium mit einer Publikation, die bei einem anderen Anbieter liegt, verknüpft wird. Eine Zusammenstellung von Quellen wird auch als „Digital Data Collection“ oder „Research Data Collection“ bezeichnet. Marilyn Deegan und Simon Tanner fassen in einem Aufsatz zur Konversion von Primärquellen die Vorteile der Digitalisierung für Geisteswissenschaftler zusammen. Dazu gehört die sogenannte „virtual reunification“, die es erlaubt, Sammlungen nach neuen Gesichtspunkten zusammenzustellen bzw. wieder zusammenzubringen, ohne dass die physischen Objekte an einem realen Ort zusammengeführt werden müssen. Aufgrund einer optimalen Erschließung können beispielsweise das Retrieval inklusive Volltext- und Multimedia-Retrieval verbessert sowie verschiedene Medienformen zusammengestellt und integriert werden.⁸⁵

3.5 Zusammenfassung

Aus den Darstellungen in diesem Kapitel zum Gegenstandsbereich der geisteswissenschaftlichen Forschung und der hermeneutischen Methode werden für diese Arbeit folgende fünf Aspekte als wesentlich erachtet:

- Die Forschungsgegenstände der Geisteswissenschaften generell und ebenso digitale Forschungsdaten werden durch Eigenschaften wie Heterogenität und Komplexität charakterisiert.
- Sie weisen darüber hinaus eine hohe Zahl an Abstraktionsstufen auf: von Sprachdaten als empirischem Material (auch Textforschung, Text Mining) bis hin zur Untersuchung von Denkmodellen.

⁸⁴ Vgl. Warwick (2004) Print Scholarship and Digital Resources, S. 379

⁸⁵ Deegan & Tanner (2004) Conversion of Primary Sources, S. 491

- Die Gegenstände und Instrumente sind zeichenbasiert, d. h. ein Zeichen nimmt auf ein anderes Bezug, auf das wiederum Bezug genommen werden kann.
- Die Verschränkung von Instrument und Forschungsgegenstand erfolgt auf semiotischer und sprachlicher Ebene.
- Auf Grundlage digitaler vernetzter Umgebungen bieten sich neue Möglichkeiten der Zusammenstellung und Verknüpfung von Forschungsdaten.

Die Auseinandersetzung mit digitalen Forschungsdaten in den Geisteswissenschaften ist keineswegs trivial und umfasst eine Vielzahl von Aspekten. Die Masterarbeit greift nur einen Teilbereich heraus, der die Tragweite der Entwicklung von e-Research und Digitalen Geisteswissenschaften sowohl hinsichtlich der technischen Möglichkeiten als auch der Komplexität erkennbar macht.

4 Digitale Forschungsdaten in der wissenschaftlichen Wertschöpfung

Das Kapitel vermittelt einen Überblick zur Verortung digitaler Forschungsdaten aus verschiedenen Perspektiven. Dazu gehören in einem ersten Schritt die Rolle der digitalen Forschungsdaten bei der zunehmend im digitalen Raum stattfindenden wissenschaftlichen Wertschöpfung sowie eine nähere Bestimmung von Informationsressourcen als Einheiten der wissenschaftlichen Wertschöpfung. Aus der wissenschaftlichen Wertschöpfung ergeben sich verschiedene Aggregationsstufen digitaler Forschungsdaten. Die Verwendung der Bezeichnungen für Forschungsdaten in den einzelnen Aggregationsstufen ist divers, da verschiedene Sichtweisen Berücksichtigung finden, weshalb sich die Begriffsbestimmung digitaler Forschungsdaten als äußerst schwierig darstellt. In einem nächsten Schritt wird „Data Curation“ als Aufgabenfeld des Datenmanagements beschrieben.

4.1 Die wissenschaftlichen Wertschöpfung in digitalen vernetzten Umgebungen

Der Begriff der wissenschaftlichen Wertschöpfung wird in dieser Arbeit für die Erzeugung wissenschaftlicher Erkenntnis durch einen Prozess mehrerer aufeinanderfolgender wissenschaftlicher Tätigkeiten verwendet. Dabei sind insbesondere zwei Tätigkeitsfelder von Interesse, die Michael Nentwich benennt: Wissensproduktion und Wissenschaftskommunikation.⁸⁶ Die Wissensproduktion reicht von der Informationsbeschaffung über die Gewinnung, Verwaltung, Verarbeitung und Analyse von Daten bis zur Aufbereitung von Forschungsergebnissen und deren Interpretation. Die Wissenschaftskommunikation spannt den Bogen von der Wissensrepräsentation über die Zusammenarbeit und den wissenschaftlichen Diskurs zum Publizieren von

Forschungsergebnissen. Zur Übersicht sind die Prozessaktivitäten in Anlehnung an Nentwich im Einzelnen aufgelistet:

Wissensproduktion:

- Beschaffung von Sekundärdaten aus vorangehenden Forschungsprozessen (Akkumulation)
- Gewinnung von Primärdaten (Experiment, Messung, Beobachtung, Quellenanalyse, Digitalisierung)
- Verwaltung von Primär- und Sekundärdaten (Ordnung, Speicherung, Auszeichnung mit Metadaten)
- Analyse (Rezeption, Analyse, Interpretieren, Vergleichen)
- Annotation, Edition, Transkription
- Gewinnung, Erstellung und Rezeption von Sekundärdaten

Wissenschaftskommunikation:

- Synthese und Präsentation der Ergebnisse (und meist Fixierung in Textform)
- Diskurs und Begutachtung
- Publikation

Der Prozess der Wissensproduktion ist iterativ und häufig durch Schleifen beschreibbar. Viele Aktivitäten überschneiden sich.⁸⁷ Bevor die Forschungsaktivitäten durch die Anwendung technischer Werkzeuge im digitalen Raum verortet werden konnten – dies wird im Fortlauf durch die Bezeichnung „traditionell“ ausgedrückt – handelte es sich in der Regel um einen singulären Prozess, der lokal an eine bestimmte Forschergruppe gebunden war. Durch die zu Beginn beschriebenen vernetzten Strukturen elektronischer Arbeitsumgebungen des e-Research und der Virtuellen Forschungsumgebungen werden einzelne Aktivitäten herausgelöst und können durch unterschiedliche Forschergruppen lokal unabhängig durchgeführt werden. Falls die Möglichkeit besteht, die

⁸⁶ Vgl. Nentwich (1999) Cyberscience, S. 5

Dienstleistungen eines Datenanbieters zu nutzen, so kann die Verwaltung von Primär- und Sekundärdaten lokal unabhängig von den anderen Aktivitäten geschehen, wenn die Daten etwa durch einen Persistent Identifier⁸⁸ referenzierbar und wieder auffindbar gemacht werden.⁸⁹ Aktivitäten der Wissensproduktion werden auf einzelne Akteure verteilt, sie sind häufig in einer Form miteinander verschränkt, so dass eine Auftrennung an mancher Stelle kaum mehr möglich ist. Dies findet in technischer bzw. medialer Form Ausdruck, so dass aus einem traditionell eher linearen Prozess ein komplexes Beziehungsgefüge entsteht. Nentwich stellt in einer Übersicht dar, welchen Einfluss die technisch-organisatorische Weiterentwicklung vom traditionellen Wissenschaftsbetrieb bis hin zur Cyberscience auf die Aktivitäten der Wissensproduktion und die wissenschaftlichen Kommunikationsprozesse hat. Dies reicht von der Datengewinnung, die im Cyberscience-Zeitalter per Modellierung und Simulation erreicht wird, bis hin zur Wissenspräsentation in Form von Hypertext-Basen.⁹⁰

Betrachtet man exemplarisch das wissenschaftliche Publizieren – bei Nentwich Teil der Wissenschaftskommunikation – so gibt es verschiedene Herangehensweisen an den Übergang vom Print- zum digitalen Paradigma des wissenschaftlichen Publizierens. Nach Stefan Gradmann ist dieser Übergang dadurch gekennzeichnet, dass sich die Zuordnung von ‚Zeigern‘ und ‚Gezeigtem‘ verändert.⁹¹ Dieser Ansatz ist für die vorliegende Arbeit von Bedeutung, da er die Einheit der wissenschaftlichen Wertschöpfungskette in ihrem strukturellen und semantischen Kontext beschreibt. Im Folgenden werden allerdings nicht allein

⁸⁷ Vgl. Nentwich (1999) Cyberscience, S. 5f.

⁸⁸ Eine Erklärung folgt in Kapitel 7.2 „Persistente Identifizierung“.

⁸⁹ Vgl. Nentwich (1999) Cyberscience, S. 5f. sowie Brase (2009) Der Digital Object Identifier (DOI), S. 60

⁹⁰ Im Anhang findet sich eine durch die Autorin erstellte Fortführung dieser Übersicht auf das Konzept des e-Research unter Berücksichtigung der für diese Arbeit relevanten Spezifika der digitalen Geisteswissenschaften – vgl. Tabelle 1 im Anhang dieser Arbeit

⁹¹ Vgl. u. a. Gradmann (2007) Some thoughts on the importance...

wissenschaftliche Publikationen als Einheiten der wissenschaftlichen Wertschöpfung behandelt, sondern auch das, was Nentwich als Primär- und Sekundärdaten benennt. Der Fokus liegt daher zunächst auf der Herausarbeitung und näheren Definition der digitalen Informationsressourcen, die die Einheiten der wissenschaftlichen Wertschöpfungskette darstellen. Im Anschluss daran knüpft sich eine Betrachtung der möglichen Aggregationsstufen digitaler Forschungsdaten.

Nancy Maron und Kirby Smith beschreiben alle die Ressourcen als wissenschaftliche Informationsressourcen („digital scholarly resources“), die beliebig einer wissenschaftlichen Fachcommunity zugeordnet werden können – sei es, dass Wissenschaftler oder eine wissenschaftliche Einrichtung im Besitz oder der Verantwortung der Daten sind oder sie für eine Wissenschaftlergemeinschaft bereitgestellt werden.⁹² Sie unterscheiden die Formen „E-only journals; Reviews; Preprints and working papers; Encyclopedias, dictionaries, and annotated content; Data; Blogs; Discussion forums; Professional and scholarly hubs“⁹³. Maron und Smith beziehen sich in ihren Ausführungen auf einen anderen Begriff von Daten, als dies in dieser Arbeit für den Bereich digitaler Forschungsdaten für Geisteswissenschaften erfolgte. Die Trennung zwischen Publikationen und Daten nach der Definition von Maron und Smith wäre angesichts der Tatsache schwierig, dass alle genannten Ressourcen in den Geisteswissenschaften Gegenstand der Forschung und damit digitale Forschungsdaten sein können, sofern man die wissenschaftlichen Informationsressourcen als Basis des wissenschaftlichen Erkenntnisprozesses versteht.

⁹² Vgl. Maron & Smith (2008) Current Models of Digital Scholarly Communication, S. 7

⁹³ Ebd.

Eine gedruckte Monographie ist eine Informationsressource der traditionellen, print-basierten wissenschaftlichen Wertschöpfung und damit ein Forschungsgegenstand der Geisteswissenschaften – unabhängig davon, ob ihr Inhalt in der wissenschaftlichen Wertschöpfung Primärquelle oder Sekundärquelle ist. Aufgrund ihrer physisch abgrenzbaren und festen Form wird sie als monolithisches und statisches Informationsobjekt bezeichnet. An dieser Stelle ist mit der Bezeichnung „Informationsobjekt“ die Einheit der wissenschaftlichen Wertschöpfung gemeint, die in verschiedenen Aggregationsstufen vorkommt.⁹⁴ In der Regel wird eine Publikation durch die wissenschaftliche Fachcommunity zur Kenntnis genommen und – typischerweise für die Geisteswissenschaften – zum Untersuchungsgegenstand einer weiteren Veröffentlichung. Da wir es im digitalen Zeitalter nicht länger allein mit Printmedien, sondern auch mit digitalen Medien zu tun haben, spielen sich die Aktivitäten der wissenschaftlichen Wertschöpfung ebenfalls im digitalen Raum ab und haben signifikante Auswirkungen auf die Bestimmung eines digitalen Informationsobjekts.

Aus dem vormals monolithischen Informationsobjekt der Publikation wird im digitalen Raum ein nun generischer Bedeutungsträger. Ein Text kann etwa mittels einer Webanwendung ausgezeichnet werden, was zur Verlagerung dieser Aktivitäten in den digitalen Raum führt. Entscheidend ist – im Falle einer Veröffentlichung in Textform – das Referenzieren von Bezugsquellen: Eine Belegstelle in einem Text muss nicht länger eine sinngemäße oder wortwörtliche Wiedergabe des Originals sein, sondern kann über die Identifizierung externer Informationsobjekte oder Teile dieser geschehen, die digital und über das WWW verfügbar vorliegen und das Originalzitat enthalten.⁹⁵ Die hypertextuelle Vernetzung ermöglicht es, jegliche referenzierbare Einheit aus dem Netz direkt

⁹⁴ Digitale Informationsobjekte und die verschiedenen Betrachtungsebenen werden im Kapitel 6 „Komplexität und Kontext digitaler Objekte“ ausführlich behandelt.

⁹⁵ Vgl. Gradmann (2007) Some thoughts on the importance..., S. 276

mit einem Dokument zu verknüpfen. Die Voraussetzung dafür ist die dauerhafte und eindeutige Adressierbarkeit.

Liegt das Informationsobjekt nicht länger in monolithischer Form vor, muss von einer Dekonstruktion des bisherigen Dokumentbegriffs als ein von seiner Umgebung formal abgrenzbares Informationsobjekt ausgegangen werden.⁹⁶ Das Dokument wird dabei zu einem buchstäblich relativen Objekt. In dieser Arbeit wird dieser Ansatz auf weitere Informationsobjekte neben der wissenschaftlichen Publikation erweitert. Daher ist nicht allein die Referenzierung von Belegstellen in Veröffentlichungen von Interesse, sondern auch – vor dem Hintergrund, dass digitale Forschungsdaten in den Geisteswissenschaften in verschiedenen Formen vorliegen – die Verknüpfung digitaler Informationsobjekte im digitalen Raum. Durch die Herstellung von Beziehungen zwischen Informationsobjekten, etwa durch Referenzen, findet sich ein Informationsobjekt in einer Informationsarchitektur wie dem WWW in einer netzartigen Struktur wieder. Wir haben es deshalb nicht nur mit der Dekonstruktion des „Dokuments“ zu tun, sondern mit einer notwendigen Verortung und einem Bestimmungsversuch des digitalen Informationsobjekts in digitalen vernetzten Umgebungen.

4.2 Aggregationsstufen digitaler Forschungsdaten

Die Aktivitäten der Wissensproduktion werden in Hinblick auf Forschungsdaten als „Aggregationsstufen“ beschrieben. Der Begriff der „Aggregation“ stammt ursprünglich aus der Chemie und beschreibt dort die Verbindung von Molekülen. In der Informationswissenschaft wird Aggregation als Metapher für kleinste Einheiten (referenzierbare Informationsobjekte, in der Chemie Moleküle) verwendet, die in einer bestimmten Weise miteinander verbunden sind. Die hier dargestellten Aggregationsstufen beziehen sich auf den Grad der Verarbeitung von

⁹⁶ Vgl. ebd. Zur Diskussion um den Dokumentbegriff siehe bspw. Buckland (1997) What is a document?

Daten in der wissenschaftlichen Wertschöpfung – angefangen bei Rohdaten bis hin zur Fixierung einer Aussage. Um bei der Metapher der Aggregation zu bleiben, lässt sich sagen, dass Forschungsdaten mit jeder Verarbeitungsstufe einen Zugewinn an Information haben, sich demnach mit einer weiteren Einheit verbinden, d. h. mit jeder Stufe entstehen – wenigstens aus konzeptueller Sicht – neue Daten. Eine Aggregation ist in diesem Sinne eine Zustandsform von Daten.

Die Aggregationsstufen unterscheiden sich von Disziplin zu Disziplin erheblich. So führen Brockman et al. zu den Quellen ('sources') in den Geisteswissenschaften aus:

„The scholars used their sources in a broad range of ways. An item could be used as "data" – a central item that is analyzed in the text. Referenced works might also be "evidence" – items used to support or show agreement for a scholar's analysis. Even an unsuccessful search may be noted as negative evidence. Sources used as exemplars were also common. Other uses of citations were to note facts (e.g., used to structure a time line), to point to items that contained further information on the topic, and to indicate the source for a framework of analysis or theory. The case-study document-analysis interviews also identified "absences" – materials that were used but not explicitly referenced or sources that the scholar would have liked to use.“⁹⁷

Das Zitat macht deutlich, dass Forschungsdaten in den Geisteswissenschaften Daten in verschiedenen Aggregationsstufen sind. Besonders im deutschsprachigen Fachdiskurs wird für die gesamte Debatte über Forschungsdaten häufig die Bezeichnung „Primärdaten“ synonym zu „Rohdaten“ oder „Forschungsdaten“ verwendet. Durch das Präfix „primär“ suggeriert „Primärdaten“ die erste Aggregationsstufe von Daten. Teilweise sind Primärdaten jedoch die aus sogenannten Rohdaten gewonnenen Erkenntnisse in aufbereiteter Form wie einer Tabelle oder eine Grafik. Auch stellt z. B. die reine Beobachtung eines Naturphänomens beispielsweise noch keine Aggregationsstufe im Sinne eines

⁹⁷ Brockman et al. (2001) *Scholarly Work in the Humanities*, S. 18

Bitstroms dar. Was würde eine Videoaufzeichnung oder die Notiz eines Wissenschaftlers zu den Beobachtungen darstellen – Rohdaten oder Primärdaten? Eine stärkere Einbeziehung der Aggregationsstufen bei der Bezeichnung digitaler Forschungsdaten ist wünschenswert, um entweder einen umfassenden Begriff wie „digitale Forschungsdaten“ oder „Artefakte des wissenschaftlichen Prozesses“⁹⁸ zu verwenden oder genauer abzugrenzen, welche Bezeichnung welcher Aggregationsstufe gilt. Im Bereich der Geisteswissenschaften ist die Datenlage auch in Bezug auf die Aggregationsstufen derart heterogen, dass eine umfassende Bezeichnung gewünscht bzw. bevorzugt wird – vor allem, wenn Forschungsdaten Disziplinen unabhängig Thema des Fachdiskurses sind.⁹⁹

Eine klare Bestimmung der Aggregationsstufen ist von Bedeutung, wenn es um die Verwaltung von Forschungsdaten und ihre Referenzierbarkeit in jedem Stadium des Lebenszyklus wissenschaftlicher Daten geht. Momentan scheint die Begriffsverwendung – zumindest solange sie nicht explizit für einen spezifischen wissenschaftlichen Arbeitsprozess bestimmt ist – willkürlich. Eine Einteilung sollte nach eindeutigen Kriterien vorgenommen werden. Es gibt mehrere Typisierungsansätze für Forschungsdaten, die sich je nach Gebrauchskontext – beispielsweise aus Sicht der Langzeitarchivierung oder nach der disziplinären Perspektive – unterscheiden. Eine Tabelle zur diesbezüglichen Übersicht findet sich im Anhang der Arbeit.¹⁰⁰

⁹⁸ Diese Bezeichnung ist ein Vorschlag von Malte Dreyer. Vgl. Kluttig (2008) Bericht über das DFG-Rundgespräch „Forschungsprimärdaten“

⁹⁹ Vgl. u. a. Klump (2009) Digitale Forschungsdaten, S. 104f.

¹⁰⁰ Vgl. Tabelle 2 im Anhang dieser Arbeit.

4.3 Data Curation

Digitale Forschungsdaten müssen in jeder Aggregationsstufe der wissenschaftlichen Wertschöpfung beschreibbar sein. Diese Aggregationsstufen werden aus Sicht des Datenmanagements durch einen Lebenszyklus von Daten („Data Life Cycle“) dargestellt, der alle Prozesse von der Datengenerierung bis zur Archivierung abbildet. Eine entsprechende Darstellung bietet die nachfolgende Abbildung.

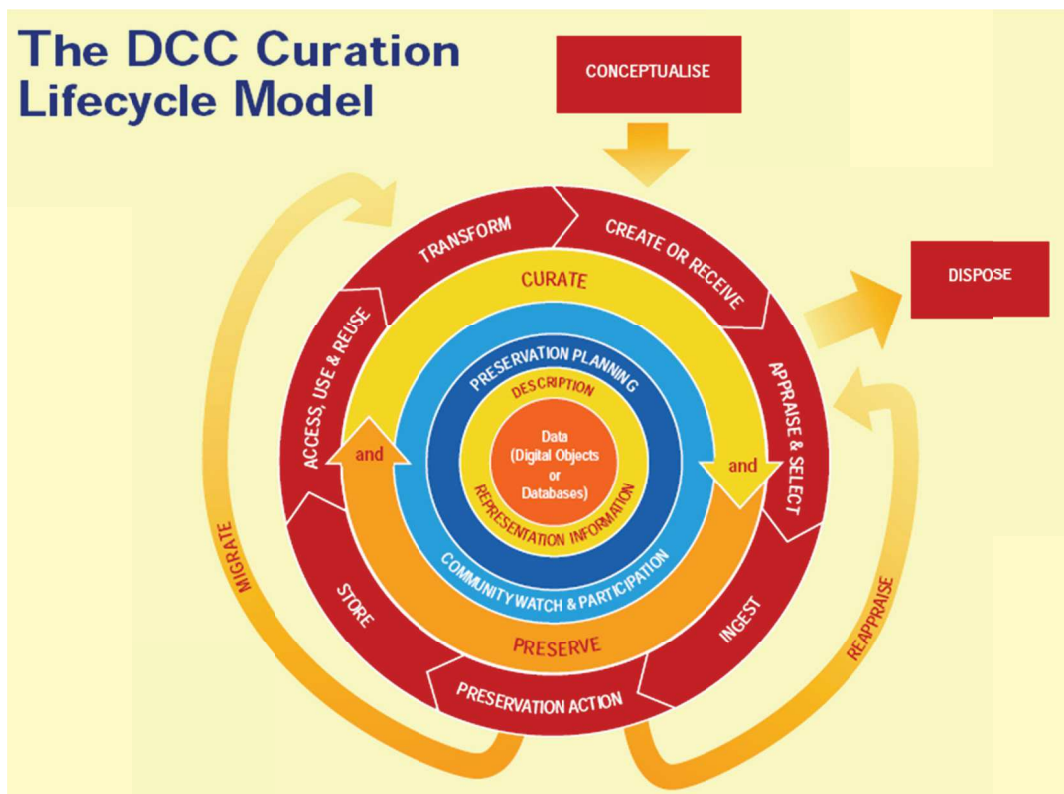


Abbildung 1: Lebenszyklus von Daten nach dem Modell des britischen Digital Curation Center¹⁰¹

¹⁰¹ <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Ein Teil des Lebenszyklus von Daten bildet die „Data Curation“, ein Konzept des Datenmanagements, das im Jahr 2000 entstanden ist und inzwischen als eine wesentliche Säule des e-Research gilt.¹⁰² Data Curation ist die Ergänzung zur „Digital Preservation“ (oder „Data Preservation“). Letztere wird vor dem Hintergrund der digitalen Langzeitarchivierung gesehen, während Data Curation am Punkt der Entstehung von Forschungsdaten ansetzt und die Datenpflege über den gesamten Data Life Cycle umfasst.

Data Curation und Digital Preservation werden unter dem Begriff der „Digital Curation“ zusammengefasst. Das britische Digital Curation Center (DCC)¹⁰³ definiert Digital Curation wie folgt:

„Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use. The digital archiving and preservation community now looks beyond the preservation, cataloguing and cross referencing of static digital objects such as documents.“¹⁰⁴

Das Zitat stellt zugleich klar, dass Digital Curation nicht länger einem statischen Objektbegriff gewidmet ist.

Neil Beagrie¹⁰⁵ richtet die Begriffsbestimmung darüber hinaus an der jeweiligen Zielgruppe und ihren Bedürfnissen aus:

„[...] The use of terms such as „archiving“, „preservation“, and „data“ can mean different things to different groups and there is often a deeply embedded local usage, which professions are reluctant to change. These difficulties have led to the recent adoption amongst some specialists of the term „digital curation“. This is a relatively new term incorporating aspects of the existing concepts „data curation“ and „digital preservation“ used

¹⁰² Vgl. Van Horik (2007) Data Curation, S. 122

¹⁰³ Vgl. <http://www.dcc.ac.uk/>

¹⁰⁴ Vgl. <http://www.dcc.ac.uk/about/what/>

¹⁰⁵ Neil Beagrie hat sowohl den Begriff „Data Curation“ geprägt als auch die sich ergänzende Sichtweise von Data Curation und Data Preservation ausgearbeitet (nach Van Horik (2007) Data Curation, S. 103f.).

primarily by the scientific and digital library communities respectively. However its use was also intended to build bridges between them and reflect new approaches.”¹⁰⁶

Zielsetzung der Digital Curation ist die Nachnutzbarkeit von digitalen Forschungsdaten für kommende Wissenschaftsgenerationen. Sie dokumentiert das vorhandene Wissen einer Disziplin oder eines Forschungsprojekts zu einem spezifischen Zeitpunkt.¹⁰⁷ Die Datenanreicherung ist dabei ein wesentlicher Aspekt:

„In most research fields, capturing ‘knowledge’ is more than just the archiving and preservation of source data and associated metadata. It generally involves interaction between creators and providers of data, the archivers of data, and most importantly the consumers of data. Successful curation of data requires data users to be able to utilise the data using their current tools and methodologies.”¹⁰⁸

Mithilfe der Digital Curation werden die Nutzer von Daten dazu befähigt, Werkzeuge und Methoden auf vorhandene Forschungsdaten, unabhängig von den Verfahren der ursprünglichen Datenerhebung, anzuwenden. Dieser Ansatz kann auf die Aussage dieser Magisterarbeit zugespielt und damit eine Parallele gezogen werden: Digitale Forschungsdaten müssen mit der entsprechenden Beschreibung ihrer strukturellen Eigenschaften angereichert werden, um eine adäquate Nutzung zu ermöglichen.

¹⁰⁶ Beagrie (2006) Digital Curation for Science, Digital Libraries, and Individuals, S. 4

¹⁰⁷ Vgl. JISC (2003) An invitation for expressions...

¹⁰⁸ Vgl. JISC (2003) An invitation for expressions...

5 Betrachtungsebenen

Aus den ersten Kapiteln dieser Arbeit kann zusammengefasst werden, dass digitale Forschungsdaten für die Geisteswissenschaften heterogen sind und aus unterschiedlichen Blickwinkeln wie der Geisteswissenschaften, der wissenschaftlichen Wertschöpfung oder der Digital Curation betrachtet werden können. Diese Vielfalt führt zu Problemen in der Begriffsbestimmung.

Ein weiteres Problemfeld besteht an der Schnittstelle verschiedener Betrachtungsebenen von Forschungsdaten: dem Verhältnis von Zeichen, Daten, Information und Wissen – und zwar

- aus Sicht der Semiotik: Wann tragen Zeichen eine Bedeutung?
- aus Sicht der wissenschaftlichen Wertschöpfung: Wann wird aus Information Wissen?
- und aus Sicht der Informatik: Wie wird Information dargestellt?

Die erste Frage wurde unter dem Aspekt der Hermeneutik bereits aufgegriffen. Letztere werden in diesem Kapitel gestellt und mögliche Betrachtungsweisen aufgezeigt. Im anschließenden Kapitel wird daran angeknüpft, indem zur Betrachtung der Vielfalt und Komplexität der Forschungsdaten ein Modell zur Hilfe genommen wird.

5.1 Zeichen, Daten, Information, Wissen

Die bereits unter Berücksichtigung der wissenschaftlichen Wertschöpfung dargestellten Aggregationsstufen sind aus einer weiteren Perspektive hochinteressant: Behandelt man beispielsweise unbearbeitete Manuskripte als „Rohdaten“, so bezieht sich die Bezeichnung „Rohdaten“ rein formal (aus syntaktischer Sicht der Informatik) auf die unbearbeiteten Daten – aus semantischer Sicht ist ein Manuskript aufgrund seiner sprachlichen Verfasstheit jedoch eindeutig Bedeutungsträger.

Betrachtet man die klassische Folge Daten → Information → Wissen, so ergeben sich in diesem Zusammenhang Fragen, die die oft angenommene Linearität der Folge deutlich relativieren: Wann sind digitale Forschungsdaten Daten? Wann wird aus ihnen Information, wann Wissen? Die folgende Fragestellung soll als exemplarischer Ausgangspunkt dienen:

Sind aus Rohdaten wie Messergebnissen abgeleitete Daten in Form einer Tabelle (aus Perspektive mancher Naturwissenschaft „Primärdaten“)

- Information, weil es sich um dekodierte Daten aus Sicht der Informatik handelt oder
- Wissen, weil Kontextwissen herangezogen werden musste, um beispielsweise die Kategorien für eine Tabelle zu bilden?

Als Ausgangspunkt für die Auseinandersetzung mit diesen Fragen werden verschiedene Betrachtungsebenen herangezogen:

Die Informatik betrachtet Daten als Bitstrom, d. h. eine Zeichenfolge¹⁰⁹, die nach bestimmten Regeln (Syntax) aufgebaut ist, um Information wie das Datenformat zu kodieren. Die Folge der Zeichen ist für den Menschen nicht verständlich. Sie wird für ihn erst unter Zuhilfenahme der Darstellung am Computer wahrnehmbar. Für den Menschen sind sie in ihrer Reinform genau genommen Rauschen. Erst unter Zuhilfenahme einer Entschlüsselungsinstanz werden die im Bitstrom kodierten Informationen überhaupt wieder in eine rezipierbare Form gebracht, d. h. dekodiert. Der Informationsbegriff meint also auf der technischen Ebene etwas anderes, als auf der Ebene der Rezeption durch den Menschen, wie sie beispielsweise bei der hermeneutischen Analyse eines Textes vorliegt. Aus semiotischer Sicht wird Information aus der Deutung von Zeichen gewonnen.

Aus Sicht der Informationstheorie nach Shannon und Weaver, die auf der kommunikationstheoretischen Betrachtung der Nachrichtenübertragung basiert, sind Daten auf physikalischer Ebene¹¹⁰ noch keine Zeichen im semiotischen Sinne, da sie keine Bedeutung tragen. Sie werden erst Zeichen, wenn sie interpretiert werden, d. h. als Information oder Rauschen bei einem Empfänger ankommen. Es ist folglich der Empfänger, der einer aus Daten bestehenden Nachricht erst die Eigenschaft „Information“ oder „Rauschen“ (oder „Redundanz“) zuweist (pragmatische Ebene). Information erweist sich demnach aus der informationstheoretischen Sicht als vom Empfänger abhängig und relativ. Festzuhalten ist, dass Information Daten voraussetzt, Daten aber nicht zwangsläufig Information sein müssen.

Das Verständnis von Daten und Information ist demzufolge vom jeweiligen Betrachtungsmodell abhängig. Im Folgenden wird dafür das „Open Archival Information System“ (OAIS) der Langzeitarchivierung herangezogen. Das OAIS ist ein ISO-Standard der digitalen Langzeitarchivierung. Es bildet ein Referenzmodell für ein Archivsystem, das notwendige Komponenten des Archivierungsprozesses und die Kommunikation der Komponenten untereinander darstellt.¹¹¹ Das OAIS-Modell beschreibt grundlegende Konzepte im Bereich der digitalen Langzeitarchivierung und legt eine Terminologie fest. An dieser Stelle ist zunächst das OAIS-Verständnis von „Information“ von Interesse: Die Grundeinheiten des OAIS-Modells sind die „Information Packages“ (IP). Der Begriff der „Information“ bezieht sich im OAIS-Modell auf jede Form von „Wissen“, das ausgetauscht werden kann. Information wird durch Daten ausgedrückt. Dabei wird ein Datenobjekt mithilfe seiner „Representation

¹⁰⁹ Zeichen werden aus Sicht der Informatik als diskrete, maschinenlesbare Informationseinheiten betrachtet.

¹¹⁰ Nach dem OAIS-Modell der Bitstream eines digitalen Objekts, Erläuterungen folgen später.

¹¹¹ Vgl. Funk (2009) Digitale Objekte und Formate

Information“ (Darstellungs- oder Repräsentationsinformation) interpretiert, wie nachfolgende Abbildung aus den OAIS-Materialien¹¹² zeigt:

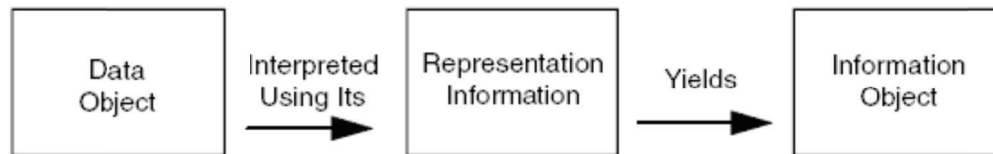


Abbildung 2: Obtaining Information from Data¹¹³

Die Darstellungsinformation beinhaltet beispielsweise Metadaten zur Sprache, in der ein Objekt verfasst ist. Informationsobjekte umfassen Datenobjekte und Metadaten, die eine pragmatische Funktion erfüllen. Wie anhand der Betrachtungsebenen eines digitalen Objekts nach dem OAIS-Verständnis gezeigt wurde, kann Information folgendermaßen festgemacht werden: logisches Datenobjekt + Metadaten (Wissen) = digitales Informationsobjekt, woraus sich ableiten lässt: Daten + Wissen = Information.¹¹⁴ Nun könnte man sich über den Bedeutungsgehalt der Metadaten streiten – das steht in Abhängigkeit zum sogenannten Wahrheitsgehalt von Information¹¹⁵, eine Diskussion die nicht Gegenstand dieser Arbeit ist. Digitale Informationsobjekte wären demnach bedeutungstragende Einheiten. Setzt man bedeutungstragend mit „Wissen“ gleich, kann hier in einer anderen Herangehensweise widersprochen werden: Erst wenn Kontextwissen angewandt wird, kann aus (bedeutungstragender) Information

¹¹² CCSDS Recommendation for OAIS – URL:
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

¹¹³ <http://public.ccsds.org/publications/archive/650x0b1.PDF>, S. 2-4

¹¹⁴ Vgl. Floridi, Luciano (2005) Is Information Meaningful Data? In: Philosophy and Phenomenological Research, Band 70, Nummer 2, S. 351-370, zitiert bei Voß (2009) Zur Neubestimmung des Dokumentbegriffs im rein Digitalen

¹¹⁵ Vgl. Voß (2009) Zur Neubestimmung des Dokumentbegriffs im rein Digitalen

Wissen werden. Die daraus abgeleitete Annahme lautet also, dass Information erst durch Kontextualisierung zu Wissen wird.

Werden digitale Informationsobjekte als geisteswissenschaftliche Forschungsgegenstände betrachtet, wie sie oben definiert wurden, so sind sie Bedeutungsträger in sprachlicher wie semantischer Hinsicht. Letztere ist folgendermaßen zu verstehen: Digitale Forschungsdaten können Wissen manifestieren (bspw. in einer Publikation). Sie repräsentieren es also. Sie sind jedoch nicht selbst Wissen. Verstanden werden sie bzw. ihr Inhalt nämlich erst durch die pragmatische Komponente der Information: D. h. sie werden als Zeichen im semiotischen Sinne interpretiert. Wird die Information interpretiert (Zeicheninterpretation), wird Kontextwissen angewandt. Die Anwendung von Kontextwissen erfolgt durch den (menschlichen) Empfänger. So entsteht Wissen.

Ausgehend von dieser Betrachtungsweise von Daten, Information und Wissen ist „Information“ nicht „Wissen in Aktion“ wie etwa bei Rainer Kuhlen.¹¹⁶ Vielmehr ist in pragmatischer Hinsicht die Information selbst in Aktion (Metadaten zu einem Datenobjekt) und wird in Form von Daten transportiert. Sie wird erst dann zu Wissen, wenn ein Akteur oder Empfänger hinzutritt, der sie in einen subjektiven Kontext einbettet. Sowohl Information wie auch Wissen lassen sich digital nur in Form von Daten/Zeichen repräsentieren. Diese Repräsentation entspricht der Qualität „Daten“, nicht der von „Information“ oder „Wissen“. Der Übergang dazu, dass ein Datenverarbeitungssystem auch Information oder Wissen erzeugen kann, steht im Zentrum des „Semantic Web“. In diesem wird über die Verknüpfungen ein digitaler Kontext erzeugt, indem weitere Verknüpfungen auf der Grundlage bestimmter programmierter Regeln automatisch erzeugt werden können. Inwiefern ein solches wissensbasiertes System tatsächlich etwas „weiß“, ist dann eine Frage für die Computerphilosophie.

5.2 Perspektivenwechsel: Beschreibung digitaler Objekte

In der gleichen Weise wie das Verständnis von Daten, Information und Wissen vom Betrachtungsmodell abhängig ist – und es zu keiner endgültigen Lösung, sondern allenfalls zu Vorschlägen kommen kann – ist die Betrachtung digitaler Forschungsdaten aus Sicht der Datenanbieter ebenso abhängig vom Betrachtungsmodell, das der Datenarchivierung und -bereitstellung zugrunde liegt. Es bestimmt, was als digitales Objekt und in welcher Art und Weise es beschrieben, prozessiert und archiviert wird. In dieser Magisterarbeit wird hier zunächst das OAIS-Modell zur Hilfe genommen, um eine Arbeitsdefinition für digitale Objekte und ihren Kontext sowie das Merkmale der Komplexität zu erhalten.

In Hinblick auf Aggregationen verteilter, heterogener digitaler Objekte wird im siebten Kapitel das paradigmatische Modell der Open Archives Initiative „Object Reuse and Exchange“ (OAI-ORE)¹¹⁷ für die Beschreibung von Webressourcen vorgestellt. Dieses stellt ein weiteres Modell der Betrachtung digitaler Objekte dar. Das ORE-Modell behandelt die Komplexität in Hinblick auf die Möglichkeiten der Verknüpfung von Webressourcen, während das OAIS-Modell generelle Aussagen zu den Ebenen der Beschreibung digitaler Objekte macht, die archiviert werden soll.

¹¹⁶ Vgl. Kuhlen (2004) Information, S. 3ff.

¹¹⁷ Vgl. <http://www.openarchives.org/ore/>

6 Komplexität und Kontext digitaler Objekte

In den vorangegangenen Abschnitten wurde der Forschungsgegenstand der Geisteswissenschaften und seine Merkmale in digitalen vernetzten Umgebungen betrachtet: Digitale Forschungsdaten können verknüpft, kollaborativ genutzt und dadurch in neuen inhaltlichen Zusammenhängen untersucht werden. Unter dieser Voraussetzung wird im aktuellen Kapitel die Perspektive gewechselt, und digitale Forschungsdaten werden aus Sicht der Digital Curation als „digitale Objekte“ betrachtet. Kenneth Thibodeau beschreibt die Vielfalt digitaler Objekte in zweierlei Hinsicht: Einerseits nimmt die Anzahl der Typen digitaler Objekte zu. Zum anderen werden sie hinsichtlich ihrer Komplexität vielfältiger.¹¹⁸ Die Typenvielfalt und die Komplexität digitaler Objekte sind zwei Aspekte, die aus Sicht der Digital Curation und – für diese Arbeit von besonderem Interesse – hinsichtlich ihrer adäquaten Darstellung und (Nach)Nutzbarkeit bewahrt werden müssen. Die Bestimmung des Begriffs „digitales Objekt“ bleibt von der zunehmenden Komplexität der Forschungsgegenstände ebenfalls nicht unberührt. Digitale Objekte bestehen nicht mehr vorrangig aus einzelnen Dateien, sondern setzen sich aus mehreren zueinander in Beziehung stehenden Teilen zusammen. Wenn digitale Objekte prozessierbar, darstellbar und nachnutzbar bereitgestellt werden sollen, ist eine adäquate strukturelle Beschreibung notwendig. Welche diesbezüglichen Anforderungen und Möglichkeiten es gibt, wird in den Kapiteln 7 „Funktionale Anforderungen an die Strukturmodellierung“ und 8 „Möglichkeiten der Strukturmodellierung“ behandelt. Das grundlegende Verständnis digitaler Objekte in dieser Arbeit wird nach einem kurzen Exkurs zu Informationsobjekten im WWW in diesem Kapitel dargestellt.

¹¹⁸ Thibodeau (2002) Overview of Technological Approaches...

6.1 Exkurs: Informationsobjekte im WWW – Hypertext und Multimedia

Der Exkurs dient als exemplarische Beschreibung des Abgrenzungsproblems von Informationsobjekten im WWW, da digitale Forschungsdaten vor dem Hintergrund ihrer Bereitstellung in digitalen vernetzten Umgebungen wie dem WWW betrachtet werden. Dabei werden die zwei wesentlichen ursächlichen Komponenten – Hypertext und Multimedia – hervorgehoben. Sie untermauern die Entstehung komplexer Beziehungsgefüge zwischen digitalen Informationsobjekten.

Hypertext meint die interne Verknüpfung von Textbausteinen und die externe Verlinkung von Dokumenten mittels sogenannter Hyperlinks. Die Fragmentierung von Text in einzelne Bausteine wird auch als Modularisierung bezeichnet. Das WWW selbst basiert auf der Idee der Verlinkung von Informationsobjekten im Internet und darf zweifelsohne als das prominenteste Beispiel für die Idee des Hypertexts bezeichnet werden. Wesentliche Merkmale des Hypertexts sind die Modularisierung und Möglichkeiten der gegenseitigen Bezugnahme von Textbausteinen und Informationsobjekten aufeinander. Sie führen zu einer Aufhebung der Linearität, wie sie von traditionellen Print-Texten bekannt ist. Hypertext führt zu einer vernetzten, nicht linearen Struktur der Information im WWW und ermöglicht den multiplen Zugang zu Informationsobjekten. Dabei kann sowohl der Leseverlauf als auch die nicht-sequentielle Anordnung von Zeichen bzw. zusammengesetzter Zeicheneinheiten als nicht-linear bezeichnet werden.¹¹⁹ Aufgrund extensiver Verlinkung von Informationsobjekten verschwinden die Grenzen zwischen ihnen.

Digitale Informationsobjekte in vernetzten Strukturen wie dem WWW werden durch „Zeiger“ identifiziert und sind mithilfe von Protokollen wie dem Hypertext

¹¹⁹ Vgl. Winko (2005) Hyper – Text – Literatur, S. 139f.

Transfer Protocol (HTTP) adressierbar.¹²⁰ Die langfristige Identifizierung von Internetobjekten ist eine der wesentlichen Herausforderungen in diesem Zusammenhang, die aber hier nicht weiter behandelt werden kann. Im Zentrum steht vielmehr die Annahme, dass basierend auf der Idee der Verknüpfung von Informationsobjekten auch die Zusammenarbeit an Forschungsobjekten im WWW vereinfacht wird.¹²¹ Das Attribut der „Diskursivität“ von Hypertext im Sinne einer gegenseitigen Bezugnahme von Ressourcen aufeinander wird in Hinblick auf die Fragestellungen dieser Arbeit im Anschluss eine Rolle spielen.

Mithilfe von Hyperlinks können Informationsobjekte verschiedener Medienformate wie Audio- und Videosequenzen oder Bilder in die Hypertexte eingebunden werden, ohne dass es dabei einer großen technischen Kompetenz bedarf. Diese Form des Medienmixes textueller Elemente mit akustischen Elementen wie Musik oder visuellen Elementen wie Bildern oder Farbgebung wird unter dem Begriff Multimedia zusammengefasst. Multimedia ist als Präsentation von Inhalten durch die Integration mindestens zweier verschiedener Medienformen wie Text, Audio, Video, Bild, Grafik und Animation gekennzeichnet.¹²² Meist ergänzen, kommentieren und unterstützen sich die verschiedenen Medienformen in einer Aggregation, und es kommt mitunter zu einem recht komplexen Zusammenspiel¹²³, dem auch in einer digitalen Repräsentation Rechnung getragen werden muss. Simone Winko stellt in dem Zusammenhang die Frage, was in einer solchen Zusammenstellung Basis der Interpretation sein kann bzw. wie sich Wort, Bild und Ton zueinander verhalten. Darstellungsform und Inhalt von über das WWW zugänglichen Dokumenten, wie beispielsweise Texten, sind eng miteinander verwoben. In der Folge kann die

¹²⁰ Vgl. Gradmann (2005) in Digitalität und Literalität, S. 303

¹²¹ Vgl. Nentwich (1999) Cybersciene, S. 18f.

¹²² Vgl. Winko (2005) Hyper – Text – Literatur, S. 139ff.

¹²³ Vgl. ebd., S. 146f.

semantische Ebene kaum mehr von der technischen getrennt werden. So kann ein Text als Forschungsgegenstand beispielsweise jeweils auf den verschiedenen Ebenen rezipiert und definiert werden.¹²⁴ Darüber hinaus erweitern sich die Möglichkeiten der freien Verlinkung inhaltlicher Bausteine durch die Entwicklung des „Mitmach-Web“ Web 2.0 (Stichwort: User Generated Content).

6.2 Betrachtungsebenen eines digitalen Objekts

Aus Sicht von Wissenschaftlern sind digitale Objekte in erster Linie *Dateien* wie Claudia Koltzenburg schreibt: „[...] on a given web site, a file is recognizable as a distinctive entity (usually given as a hyperlink). The boundary of such an entity is marked by a kind of packaging indicator made up of a file name, a boundary sign, and a file format abbreviation [...]”¹²⁵.

Eine generische Definition besagt, dass alles als digitales Objekt betrachtet werden kann, was digital vorliegt, gleich welchen Typs oder Format: eine Grafik, ein Bild, ein Textdokument, eine Videodatei, eine Audiodatei, ein Anwendungsprogramm wie z. B. ein Internetbrowser oder auch eine Datenbank. Nach Kenneth Thibodeau¹²⁶ können alle digitalen Objekte durch die gemeinsame Eigenschaft der Mehrfachvererbung charakterisiert werden. Ihre Eigenschaften werden durch drei Klassen vererbt. Jedes digitale Objekt ist zugleich physisches, logisches und konzeptuelles Objekt mit jeweils spezifischen Eigenschaften.

Als „physical object“ besteht es aus kodierten Zeichen (Zeichen nach informatischen, nicht semiotischen Verständnis) auf einem Trägermedium. Als „logical object“ besteht es als prozessierbare Einheit und damit als das Objekt, das von einer Anwendungssoftware verstanden wird. Das Verständnis von Logik

¹²⁴ Vgl. Hiebler (2005) Von der Medienkulturgeschichte digitaler Codierungen ..., S. 85

¹²⁵ Koltzenburg (2008) Check-Listing digital objects in context, S. 229

¹²⁶ Vgl. Thibodeau (2002) Overview of Technological Approaches...

entstammt in dem Fall der „Sichtweise“ einer Anwendungssoftware. Logische Objekte können andere logische Objekte enthalten. Jedes logische Objekt kann identifiziert werden. Als „conceptual object“ besteht das Objekt als bedeutungstragende Einheit, die Menschen als Information wahrnehmen (z. B. ein Buch mit Abbildungen). Die Eigenschaften konzeptueller Objekte sind signifikant in der Realwelt. Die Struktur des konzeptuellen Objekts steht in direktem Zusammenhang mit der Struktur des logischen Objekts. Da aber jeder konzeptuelle Inhalt in verschiedener Form digital kodiert sein kann, können sich beide Strukturen dennoch voneinander unterscheiden. So kann der Inhalt eines digitalen Dokuments beispielsweise in Form eines PDF-Dokumentes oder auch als Bild-Datei kodiert sein. Das bedeutet, dass ein konzeptuelles Objekt verschiedene digitale Repräsentationen haben kann. Letztlich kann dies auch auf komplexe Objekte ausgeweitet werden. Das Ziel einer maschinenlesbaren Darstellung bleibt bestehen.¹²⁷

Weiterhin gibt Thibodeau die Verhältnisse der einzelnen Betrachtungsebenen zueinander durch die Typen „one-to-one“, „one-to-many“, „many-to-one“ und „many-to-many“ an. Ein konzeptuelles Objekt kann vier logische Objekte enthalten, was z. B. eine „one-to-many-Relation“ ist. Eine Webseite und ihre Hyperlinks stellen eine „many-to-many“-Relation dar. Für eine korrekte Prozessierung und Archivierung der digitalen Objekte müssen diese Verhältnisse explizit gemacht werden.¹²⁸

6.3 Beschreibungsebenen digitaler Objekte im OAIS-Modell

Da sich die Masterarbeit auf den Ansatz und die Terminologie des OAIS stützt, werden im Anschluss die Beschreibungsebenen digitaler Objekte im OAIS-Modell näher beschrieben. In Anlehnung an Thibodeau wird das digitale Objekt

¹²⁷ Vgl. Thibodeau (2002) Overview of Technological Approaches...

¹²⁸ Vgl. ebd.

im OAIS-Modell als eine Menge von Bit-Sequenzen verstanden, die sich auf physischer, logischer und konzeptueller Ebene beschreiben lassen. Unter den Gesichtspunkten der Aufbereitung, Archivierung, Bereitstellung und nachhaltigen Verfügbarkeit können mithilfe der Betrachtungsebenen Aussagen über die Konstitution und Beschaffenheit digitaler Objekte gemacht werden.¹²⁹

Das physikalische Objekt wird auf einer Datenebene beschrieben. Das logische Datenobjekt wird auf einer Interpretationsebene beschrieben. Interpretation bezieht sich hier auf die Formatinformation. Das konzeptuelle Objekt deckt sich mit der Annahme über digitale Objekte als Forschungsdaten (Bedeutungsebene): Es kann sich um ein Buch, ein Gemälde, eine Sprachaufnahme handeln, oder wie Funk beschreibt, die Gesamtheit der Funktionalität für den Nutzer unter Zuhilfenahme von Soft- und Hardware wie beispielsweise das Durchsuchen einer Datenbank und das Editieren einer Textdatei sein. Das konzeptuelle Objekt wird auch als Informationsobjekt bezeichnet.¹³⁰

6.4 Komplexe digitale Objekte

Abgesehen von den Betrachtungsebenen eines digitalen Objekts, wie es das OAIS-Modell liefert, ist die Begriffsbestimmung eines „digitalen Objekts“ und die Bestimmung seiner Konstitution schwierig, wie folgendes Zitat belegt:

„The characteristics of the scientific data objects determine to a large extent the required data curation activities. [...] One of the problems here relates to the nature of digital objects. A traditional document, like a book, tends to be correspondent to a discrete physical object. On the other hand a digital object can be unstable, dynamic and without boundaries.“¹³¹

¹²⁹ Vgl. Funk (2009) Digitale Objekte und Formate

¹³⁰ Vgl. Funk (2009) Digitale Objekte und Formate

¹³¹ Van Horik (2007) Data Curation, S. 134

Mit dieser Aussage ist ein wesentliches Problem der Beschreibung digitaler Objekte benannt: Wodurch werden die Grenzen eines digitalen Objekts konstituiert? Dies scheint insbesondere im Vergleich zur Abgeschlossenheit physisch vorhandener Objekte zum Tragen zu kommen.

Komplexe digitale Objekte bestehen aus Aggregationen digitaler Objekte, die miteinander verknüpft sind. Die Komplexität wird durch verschiedene Eigenschaften beschrieben¹³²:

- Komponiertheit: Digitale Objekte können sich aus anderen Objekten zusammensetzen. Welches Objekt als konzeptionelles Objekt angesehen wird, bleibt Sache des Betrachters.
- Granularität: Diese Eigenschaft bezieht sich auf das Verhältnis der Objektteile zwischen konzeptioneller und logischer Ebene. Hier muss kein 1:1-Verhältnis vorliegen. Ebenso kann es vorkommen, dass ein konzeptionelles Objekt aus mehreren Datenobjekten besteht wie beispielsweise eine Webseite oder Zeitschriften, die einzelne Artikel enthalten, die wiederum aus Seiten aufgebaut sind usw. In diesem Sinne können einzelne Granulationsniveaus auch verschiedenen Genres bzw. Dokumenttypen entsprechen.
- Verteilung: Hier greift eine Problematik, die besonders im Kontext dieser Arbeit von Interesse ist: Die lokale Bestimmung verteilter Datenobjekte ist schwierig.
- Dynamik: Die Objekte können beispielsweise von Nutzern verändert werden. Ludwig beschreibt weitere Eigenschaften wie Aktivität von Objekten (Software), komplexe inhaltliche Strukturierung von Objekten (Hypertext)

¹³² Vgl. Ludwig (2008) Digitale Objekte und ihre Eigenschaften sowie eigene Darstellung

und Nutzungsrechte als Eigenschaften, die Objekten „von außen“ zugeschrieben werden.¹³³

- Atomarität: Wie lässt sich ein digitales Objekt als diskretes Ganzes beschreiben? Was ist seine minimal mögliche Erscheinungsform und was ist in diesem Zusammenhang eine atomare Einheit? In dieser Hinsicht bereitet beispielsweise die Tatsache Probleme, dass jedes einem Objekt zugeordnete Objekt wieder ein eigenes Objekt sein kann und sich damit schwer eine Einheit finden lässt.
- Distinktivität: Wie lässt sich ein Objekt von einem anderen unterscheiden? Auch digitale Forschungsdaten existieren physisch als Bitstrom. In einer digitalen Welt besteht allerdings auch seine Umgebung physisch aus Bitströmen, so dass hier eine Distinktion – zumindest auf physischer Ebene – schwer ist.

Hilfreich ist in diesem Zusammenhang die strukturelle Beschreibung des Kontextes digitaler Objekte, wie er nachfolgend dargestellt wird.

6.5 Strukturelle Merkmale und Beschreibung im OAIS-Modell

Unter der „strukturellen Beschreibung“ wird die Anreicherung digitaler Objekte oder Aggregationen digitaler Objekt mit objektspezifischen Metadaten und der Darstellung der Beziehungen zueinander verstanden (Kontextualisierung). Die strukturelle Beschreibung bildet die Schnittstelle zwischen logischer und konzeptueller Ebene. Ziel der Langzeitarchivierung ist, ein digitales Objekt auf der konzeptuellen Ebene nutzbar zu machen. Dabei gilt es die wesentlichen Eigenschaften („significant properties“) konzeptioneller Objekte zu bewahren und digital adäquat abzubilden. Zu diesen Eigenschaften gehören der semantische Kontext und die Struktur, die sich auf das Format eines logischen Datenobjekts

¹³³ Vgl. Ludwig (2008) Digitale Objekte und ihre Eigenschaften

bezieht. Die strukturelle Beschreibung stützt sich auf die logischen Datenobjekte und wird mit der „Representation Information“ (Darstellungsinformation) dargestellt. Sie dient der Interpretation des Datenstroms und gewährt damit den Zugang zum Datenobjekt. Datenobjekt und Darstellungsinformation bilden zusammen das Informationsobjekt wie es in der Abbildung 3 dargestellt ist.

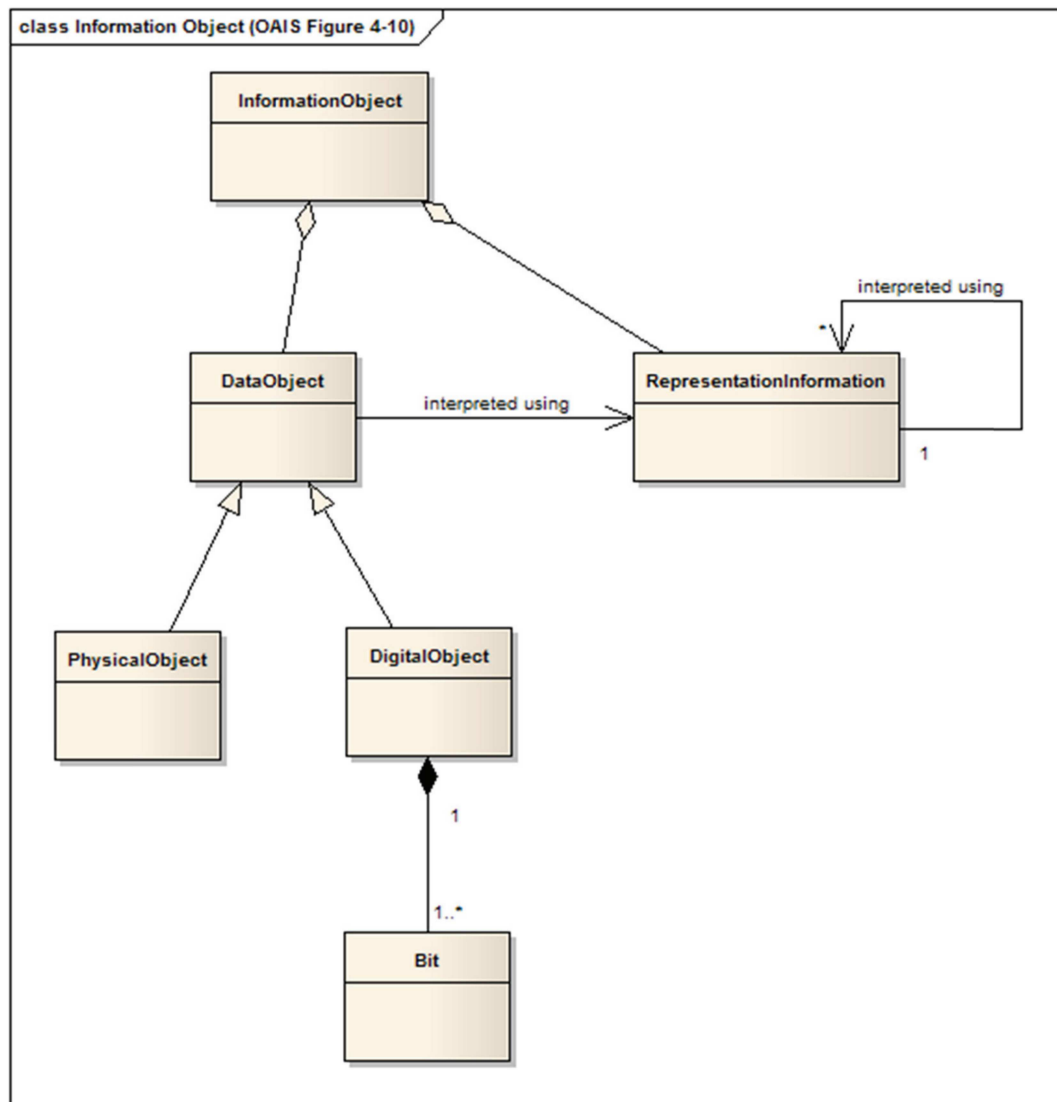


Abbildung 3: Das digitale Informationsobjekt im OAIS Reference Model¹³⁴

Jeder Prozess im OAIS-Modell basiert auf der Transformation einer konkreten Einheit – dem sogenannten „Information Package“ (IP) als zentrales Element. Das „Information Package“ ist ein konzeptueller Container bestehend aus der „Content Information“, die das digitale Informationsobjekt darstellt, wenn es als die zentrale Archivierungseinheit im OAIS-Modell betrachtet wird. Aus Sicht

traditioneller Archive ist Content Information ein Synonym für Record (archivarische Bezugseinheit)¹³⁵, das von Nutzern mit einem bestimmten Kontextwissen wahrgenommen wird. Diese „Zielgruppe“ wird „designated community“ genannt. Für die vorliegende Arbeit wird folgende Arbeitsdefinition für das Verhältnis der Begriffe „digitales Objekt“ und „digitale Forschungsdaten“ generiert: Digitale Objekte sind digitale Informationsobjekte nach dem OAIS-Vokabular. Digitale Forschungsdaten sind nach selbigem die „Content Information“ für die jeweilige Wissenschaftscommunity.

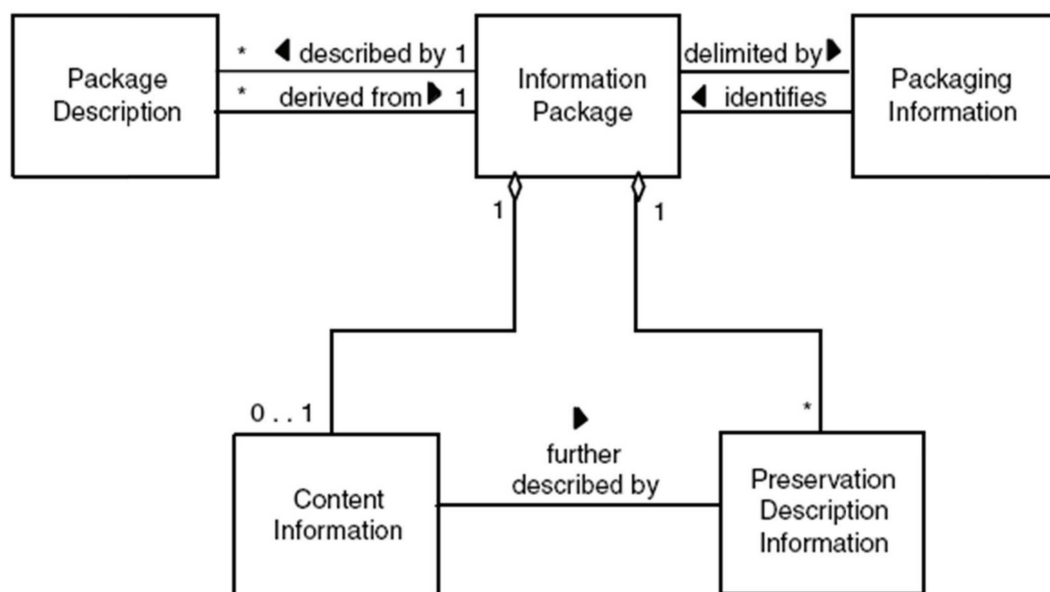


Abbildung 4: Das Konzept des Information Packages im OAIS Referenz-Modell¹³⁶

Daneben besteht ein IP aus der sogenannten „Preservation Description Information“. Es ist das Subjekt im OAIS-Modell. Ein IP wird gekapselt identifiziert über die „Descriptive Packaging Information“.¹³⁷ Die strukturelle

¹³⁴ <http://isegserv.itd.rl.ac.uk/blogs/alistair/wp-content/uploads/2007/10/InformationObject.png>

¹³⁵ Vgl. CCSDS (2002) OAIS Reference Model, S. B-2

¹³⁶ http://thomas.kiehnefamily.us/thomas_files/inf389k-final_assignment_html_7cc6006a.png

¹³⁷ Vgl. CCSDS (2002) OAIS Reference Model, S. 2-4ff.

Beschreibung, wie sie in dieser Arbeit verwendet wird, bezieht sich auf die Ebene der „Preservation Description Information“ im OAIS-Modell.

Für diese Arbeit ist die Ebene der „Preservation Description Information“ (PDI) als Datenbeschreibung für die Archivierungsprozesse von besonderem Interesse, da sie die relevanten Informationen über die Struktur und den Kontext digitaler Objekte enthält. Hier ist zu beachten, dass nicht die formale Struktur wie oben gemeint ist (die in Bezug auf das logische Datenobjekt in der Darstellungspräsentation festgehalten wird), sondern sich Struktur auf eine konzeptuelle Ebene bezieht.

Die PDI steht in engem Zusammenhang mit der Content Information und baut auf ihr auf. Sie fasst folgende Informationen zusammen: Provenance, Context, Reference und Fixity.¹³⁸

¹³⁸ Vgl. CCSDS (2002) OAIS Reference Model, S. 2-4ff.

Content Information Type	Reference	Provenance	Context	Fixity
Space Science Data	<ul style="list-style-type: none"> • Object identifier • Journal reference • Mission, instrument, title, attribute set 	<ul style="list-style-type: none"> • Instrument description • Processing history • Sensor description • Instrument • Instrument mode • Decommulation map • Software interface specification 	<ul style="list-style-type: none"> • Calibration history • Related data sets • Mission • Funding history 	<ul style="list-style-type: none"> • CRC • Checksum • Reed-Solomon coding
Digital Library Collections	<ul style="list-style-type: none"> • Bibliographic description • Persistent identifier 	<ul style="list-style-type: none"> • For scanned collections: <ul style="list-style-type: none"> • metadata about the digitisation process • pointer to master version • For born-digital publications: <ul style="list-style-type: none"> • pointer to the digital original • Metadata about the preservation process: <ul style="list-style-type: none"> • pointers to earlier versions of the collection item • change history 	<ul style="list-style-type: none"> • Pointers to related documents in original environment at the time of publication 	<ul style="list-style-type: none"> • Digital signature • Checksum • Authenticity indicator
Software Package	<ul style="list-style-type: none"> • Name • Author/Originator • Version number • Serial number 	<ul style="list-style-type: none"> • Revision history • License holder • Registration • Copyright 	<ul style="list-style-type: none"> • Help file • User guide • Related software • Language 	<ul style="list-style-type: none"> • Certificate • Checksum • Encryption • CRC

Abbildung 5: Beispiele von PDI-Typen im OAIS Referenz-Modell¹³⁹

Nachfolgend werden die vier genannten Informationstypen näher beschrieben. Sie finden sich in Teilen in den funktionalen Anforderungen des Anschlusskapitels wieder. Sie dienen der Herausarbeitung der strukturellen Merkmale digitaler Objekte. Explizite Aspekte der Langzeitarchivierung spielen in der Beschreibung keine Rolle.

„Reference“ meint neben der natürlichsprachlichen inhaltlichen Beschreibung des Kontexts die technische Identifikationsinformation. Auf der Basis der

¹³⁹ <http://www.payer.de/digitalebibliothek/digibib0206.gif>

Identifizierung von „Content Information“ können Referenzen und Beziehungen zwischen Datenobjekten ausgedrückt werden. Sie ist damit Voraussetzung für die Kontextinformation. In Anlehnung an die semiotische Verfasstheit von Forschungsgegenständen wird diese Information benötigt, um das Verhältnis zwischen „Zeiger“ und „Gezeigtem“ zu formalisieren.

„Provenance“ meint die Provenienzinformation. Sie bezieht sich auf den Entstehungs- und Herkunftsrahmen der „Content Information“. Es werden demnach die Ereignisse der Erzeugung, der Aufnahme, der Erhaltung und rechtliche Statusänderungen festgehalten. Darunter fällt beispielsweise die Dokumentation eines Erhaltungsereignisses wie die Migration von Datenformaten, die zu Manifestationen eines Datenobjekts werden (siehe „Context“).

„Context“ bezieht sich auf die Darstellung der Relationen zu weiterer Information außerhalb eines „Information Packages“. Die Kontextinformation wird genutzt, um die Entstehung bzw. den Grund der Entstehung der „Content Information“ zu dokumentieren und die technische und inhaltliche Beziehung zu weiterer „Content Information“ festzuhalten. Die inhaltlichen Beziehungen können zum einen durch inhaltsbezogene Kontextinformationen, sowie zum anderen durch die Darstellung von Manifestationen derselben „Content Information“ ausgedrückt werden. So kann es beispielsweise verschiedene Versionen eines Datenobjekts geben. Wird etwa das Format eines „Archival Information Packages“ (AIP) migriert, so wird eine neue Manifestation in Form eines AIP¹⁴⁰ erstellt, das diesen Vorgang in seiner Kontextinformation festhält.

„Fixity“ meint die Integrität (Unversehrtheit) des digitalen Objekts, die z. B. mit Hilfe einer Prüfsumme bewahrt wird. Sie bezieht sich auf den Bitstrom des

digitalen Informationsobjekts. Die Integritätsinformation beinhaltet die Prüfsumme selbst, die Information zur Berechnungsfunktion sowie den Bezug zur Prüfsumme. Die Integrität der PDI selbst ist nicht überprüfbar. Integrität schützt die Content Information vor undokumentierten bzw. unberechtigten Änderungen.¹⁴¹

6.6 Zusammenfassung

Der Kontext eines digitalen Objekts ist entscheidend für die Darstellung der Komplexität digitaler Objekte. Die Darstellung des Kontexts ist nach dem Verständnis des OAIS-Modells Bestandteil der „Preservation Description Information“. Er wird mittels Relationen digitaler Objekte dargestellt. Bei der Prozessierung von Datenobjekten müssen die Kontextinformationen mitgeliefert werden, um das konzeptuelle Objekt adäquat darzustellen.

¹⁴⁰ Die Begriffsbestimmung des Archival Information Packages ist hier nicht entscheidend. Daher sei auf das Referenzmodell verwiesen: CCSDS (2002) OAIS Reference Model, S. 2-4ff.

¹⁴¹ Vgl. CCSDS (2002) OAIS Reference Model, S. 2-4ff. sowie Margulies (2009) Digitale Daten als Quelle der Geschichtswissenschaft, S. 385ff.

7 Funktionale Anforderungen an die Strukturmodellierung

In diesem Kapitel werden exemplarische funktionale Anforderungen formuliert, die sich aus dem Umgang mit digitalen Forschungsdaten in der hermeneutisch geprägten Wissenschaftspraxis ergeben. Die Hermeneutik zielt auf die Interpretation von Zeichen, d. h. Forschungsgegenstände werden auf ihre Bedeutung hin untersucht. Die Bedeutungsfindung geschieht beispielsweise durch den Vergleich von Forschungsgegenständen oder durch die genauere Bestimmung des inhaltlichen wie formalen Kontexts, in den ein Forschungsgegenstand eingeordnet werden kann.

Im Allgemeinen sind Funktionen Bestandteile eines Prozesses, so wie die Sprache eine Funktion im Kommunikationsprozess übernimmt. Dabei wird eine Funktion als Aktivität und damit in ihrer ursprünglichen, nicht-mathematischen Verwendung definiert.¹⁴² Das abgeleitete Adjektiv „funktional“ bedeutet, dass eine Aktivität beispielsweise in einem bestimmten Forschungsprozess berücksichtigt wird. „Funktional“ ist nicht im harten Sinne der Softwaretechnik zu verstehen, sondern bezieht sich auf die Umsetzung eines methodischen Ansatzes. Für die generelle Sicht auf methodische Gemeinsamkeiten werden mehrere Literaturquellen als Arbeitsgrundlage herangezogen.

Das Kapitel und die Masterarbeit an sich haben über die Formulierung exemplarischer funktionaler Anforderungen hinaus keinen Anforderungskatalog oder eine Spezifikation im Sinne technischer Vorgaben zum Ziel.¹⁴³

¹⁴² Vgl. Mittelstraß (2005) Enzyklopädie, S. 590, Stichwort: Funktion

¹⁴³ Vgl. als weiteren Anhaltspunkt für die Formulierung funktionaler Anforderungen u. a. Verhaar (2009) Enhanced Publications: Object Models and Functionalities

7.1 Diskursivität und Bezugnahme

„The processes of reading and searching, developing context, and rereading and researching are at the heart of humanities scholarship.“¹⁴⁴

Wie das Zitat veranschaulicht, gibt es grundlegende Annahmen über die Wissenschaftspraxis in den Geisteswissenschaften, die zur Bedeutungsfindung und Wissensproduktion herangezogen werden. Im wissenschaftlichen Erkenntnisprozess der Literaturwissenschaft beispielsweise dominieren die Tätigkeiten des Beschreibens, Analysierens und Interpretierens von Forschungsquellen. Es werden die sprachlichen und formalen Merkmale eines Textes beschrieben. Die Analyse von Texten stützt sich auf ihre Struktur und deren Beschreibung und stellt semantische Beziehungen zwischen Textelementen her. Im Schritt der Interpretation folgt die (begründete) Bedeutungszuweisung. In der Literaturwissenschaft besteht – wie in den meisten anderen (geisteswissenschaftlichen) Disziplinen – keine Standardisierung hinsichtlich der Form des wissenschaftlichen Vorgehens.¹⁴⁵ Was das Kriterium „wissenschaftlich“ erfüllt, ist variabler als z.B. in den methodisch enger gefassten Naturwissenschaften.

Im Zentrum steht demzufolge die gegenseitige Bezugnahme von Forschungsgegenständen aufeinander – sowohl aus semiotischer Sicht (die Zuweisung von Bedeutung durch Zeichen), als auch vor dem Hintergrund der Verknüpfungsmöglichkeiten im digitalen Raum.

Geisteswissenschaften sind selbstreflexiv, weil sie sich im Prozess der Bedeutungsfindung und Wissensproduktion auf sich selbst beziehen – das heißt auf die Instrumente und Gegenstände der Forschung. Die gegenseitige Bezugnahme der Forschungsgegenstände erfolgt im digitalen Raum aus

¹⁴⁴ Brockman et al. (2001) *Scholarly Work in the Humanities*, S. 4

¹⁴⁵ Winko (2005) *Hyper – Text – Literatur*, S. 150

technischer Sicht durch die Herstellung von Verknüpfungen zwischen digitalen Objekten. Reflexivität kann dabei als eine Eigenschaft dieser Verknüpfungen verstanden werden. Ein Gegen-stand befindet sich somit potentiell auch in einer Relation zu sich selbst.

Der diskursive Charakter der Geisteswissenschaften äußert sich sowohl im Bereich der Wissensproduktion (Bedeutungsfindung durch die Interpretation von Zeichen), als auch in der wissenschaftlichen Kommunikation (Bezugnahme von Publikationen aufeinander ausgedrückt durch Referenzen). Nimmt man ein Kontinuum der wissenschaftlichen Wertschöpfung an, gibt es kein Ende des Diskurses, da die Instrumente und die Gegenstände der Forschung im ständigen Austausch und einer Verschränkung ihrer Eigenschaften miteinander stehen.

Ein weiterer Aspekt, und zugleich die Ursache für die Popularität und die Notwendigkeit des Themas Forschungsdaten innerhalb der Fachdiskurses, ist, dass durch die Bereitstellung – vor allem nach einem freien Zugangsmodell – die konkreten wissenschaftlichen Erkenntnisprozesse nachvollziehbar gemacht werden. Dies kann durch die Explikation der Bezüge zwischen Forschungsgegenständen und Instrumenten erfolgen.

7.2 Prinzipien im Umgang mit Forschungsdaten

John Unsworth identifiziert in einem Paper mit dem Titel „Scholarly Primitives: what methods do humanities researchers have in common and how might our tools reflect this?“ sieben Nutzungsszenarien als grundlegend im Umgang mit Forschungsdaten bzw. Forschungsquellen in der Wissenschaft¹⁴⁶: „Discovering, Annotating, Comparing, Referring, Sampling, Illustrating, Representing“. Er beschreibt sie als innerhalb aller Wissenschaftsdisziplinen verbreitete und über die Zeit stabile Funktionen, die in keiner Abhängigkeit zu der theoretischen

¹⁴⁶ Vgl. Unsworth (2000) Scholarly Primitives

Ausrichtung der Forschung oder einem Medientyp stehen. Sie sind als funktionale Ansätze zu sehen, die der eigentlichen Methodik einzelner Disziplinen vorausgehen und damit als Ausgangspunkt für die Anwendung verschiedener Zugangsinstrumente und -methoden gelten können. Unsworth stellt das *Referring* und *Representing* u. a. im Kontext digitaler vernetzter Information als grundlegend für die anderen genannten Ansätze heraus.¹⁴⁷ Die nachfolgenden Anforderungen stützen sich im Wesentlichen auf Unsworth.

7.3 Referring: Verweise

Prinzipiell stellt ein Verweis eine Relation zwischen digitalen Objekten her. Mit ihm wird eine Eigenschaft zwischen den Gegenständen dar- und hergestellt. Relationen sind beispielsweise *A steht neben B* (Nähe) oder *A referenziert B* (Bezug).

Die Zitation ist eine Form des Verweises auf externe Objekte. Aus historischer Perspektive wird die Referenzierbarkeit von Textstellen erst durch kleinere, möglichst exakt abgrenzbare Informationseinheiten überhaupt möglich. So war es die Codex-Form des Buches, die mit der Möglichkeit der Seitenzählung auch einen entsprechend präzisen Verweis zuließ. Die Form der Zitation wird nun auf komplexe digitale Objekte ausgeweitet. Wissenschaftler können in ihren Publikationen nicht nur auf andere Publikationen verweisen, sondern darüber hinaus eine direkte Verbindung zum Untersuchungsgegenstand herstellen. Die Forschungsdaten werden also direkt in den Publikationskreislauf eingebunden.

Die Relationen zwischen digitalen Objekten („Linking“) können näher spezifiziert werden. Dies umfasst den Relationstyp, die formale bzw. technische Umsetzung und den Parameter der Dauerhaftigkeit (dazu siehe unter dem Punkt Identifizierung mehr). Mithilfe von Relationen zwischen digitalen Objekten wird

¹⁴⁷ Vgl. Unsworth (2000) Scholarly Primitives

ihr Kontext generiert. Es entsteht ein logischer Zusammenhang zwischen digitalen Objekten. Eine konkrete Spezifizierung muss in gemeinsamer Arbeit mit Wissenschaftlern bzw. Benutzern digitaler Forschungsdaten vorgenommen werden. Im Folgenden werden daher einige exemplarische Relationstypen vorgestellt.

7.3.1 Comparison: Vergleiche

Nach John Unsworth ist der Vergleich von Quellen ein grundlegender methodischer Ansatz in den Geisteswissenschaften:

“Scholars in many different disciplines, working with many different kinds of materials, want to compare several (sometimes many) objects of analysis, whether those objects are texts, images, films, or any other species of human production.”¹⁴⁸

Um die digitalen Objekte auf inhaltlicher Ebene im Sinne der Zeicheninterpretation vergleichen zu können, muss klar sein, ob und in welcher Relation sie zueinander stehen. Dies ist zum einen direkt für die Wissensproduktion relevant. Zum anderen spielt es für die Nachvollziehbarkeit im Prozess der wissenschaftlichen Kommunikation bzw. des Diskurses eine grundsätzliche Rolle. So sollte etwa die Zugehörigkeit möglicher Vergleichstexte (Sekundärtexte) zu einer Quelle aufgezeigt werden können. Das bedeutet konkret, dass für diese Form der Bezugnahme bestimmte Relationstypen vorhanden sein sollten, die verschiedene Versionstypen eines Objekts innerhalb einer Aggregation abbilden können.

7.3.2 Provenienz

¹⁴⁸ Vgl. Unsworth (2000) Scholarly Primitives

Auch die Provenienz (Originalherkunft) eines kulturellen Artefakts lässt sich mithilfe von Relationen ausdrücken, indem etwa ein Bezug zur originären Sammlung hergestellt wird. Die Provenienz eines Forschungsgegenstands ist entscheidend für seine Einordnung in den Forschungskontext wie etwa Werner Schweibenz sie für Museumsobjekte beschreibt als: „[...] die originale Gesellschaft, die ein Objekt hervorgebracht hat. Diese Gesellschaft gibt den Kontext vor, für den das Objekt geschaffen und in dem es verwendet wurde [...]“.¹⁴⁹

7.3.3 Annotation, Edition, Transkription

Die Begriffe Annotation, Edition und Transkription beziehen sich auf die Analyse und Präsentation von Forschungsdaten. Sie können unter anderem als Instrument der Bearbeitung von Forschungsdaten und als Forschungsgegenstände (etwa im Sinne von Sekundärquellen) verstanden werden.

Kann eine Annotation, Transkription oder Edition selbst als digitales Objekt identifiziert werden, so sollte die strukturelle Beschreibung die Herstellung von reflexiven Verknüpfungen zwischen Originalobjekt und den Bearbeitungen ermöglichen.

Annotation meint die Anreicherung von Korpora mit Kommentaren. Dabei muss ein Bezug der Annotationen auf kleinere Einheiten als den Dokumenteinheiten möglich sein. Wie granular dieses Bezugsraster ausfällt, ist jeweils des Forschungsanliegen und der Praktikabilität zu prüfen. Prinzipiell ist ein Bezug bis auf Lemma-Ebene oder Zeichenebene denkbar, was z. B. in der Philologie oder der Linguistik relevant ist. Technisch erfolgen Auszeichnung und Beschreibung dieser digitalen Objekte u. a. auf Basis von Auszeichnungssprachen wie XML bzw. XML-basierten Standards wie TEI.

¹⁴⁹ Schweibenz (2008) Vom traditionellen zum virtuellen Museum, S. 108f.

Die Transkription meint aus sprachwissenschaftlicher Perspektive die Übertragung von einem Phonem einer Schrift in das entsprechende Phonem einer anderen Schrift (z. B. vom griechischen in das lateinische Alphabet) oder die Niederschrift gesprochener Sprache. Editionswissenschaftler etwa transkribieren ganze Texte phonemgetreu, d. h. buchstabengenau. Im Bereich der Werkzeuge für Geisteswissenschaftler wird dies meist durch sogenannte Editoren unterstützt, die neben einer Transkription auch die Auszeichnung der Originalobjekte und der Transkriptionen ermöglichen. Dies ist vergleichbar mit einer Volltexterfassung digitaler Texte, wobei häufig auf den Standard TEI zurückgegriffen wird.

Editionen sind zum einen Ausgaben im Sinne von Auflagen eines Werkes (z. B. Buch, Faksimile) und zum anderen inhaltlich oder formal aufbereitete Ausgaben eines Werkes. Der Editions begriff unterscheidet sich darüber hinaus zwischen den einzelnen geisteswissenschaftlichen Disziplinen. Historiker editieren Quellen, um sie archivalisch zu erschließen, während Philologen sogenannte „kritische Editionen“ anfertigen, um Texte in ihrem Kontext darzustellen und einem besseren Vergleich bzw. der Kritik anhand anderer Quellen zu unterziehen. Für das Anfertigen kritischer Editionen sind Verweise und die Bezugnahme auf Forschungsquellen wesentliche Instrumente des Forschungsprozesses.

7.4 Persistente Identifizierung

Digitale Forschungsdaten müssen in vernetzten Umgebungen wie dem WWW dauerhaft und eindeutig identifiziert werden können, um referenzierbar zu sein. Die Identifizierung von Forschungsdaten ist auch in Hinblick auf die Erkennbarkeit einer Aggregation und der einzelnen Komponenten einer Aggregation relevant.

Die Identifizierung von Informationsobjekten ist eineindeutig. Dabei geht es nicht mehr um die lokale Identifizierung etwa durch Hyperlinks, sondern um die identitätsorientierte Referenzierung. Wie digitale Objekte als Forschungsdaten in allen Stadien ihres Lebenszyklus beschrieben werden können¹⁵⁰, steht in Abhängigkeit zur Bestimmung der Aggregationsstufen. Die Identifizierung digitaler Objekte sollte dauerhaft sein und wird mittels sogenannter Persistent Identifier ermöglicht. Persistent Identifier sind konstant gültig und global auflösbar. Es gibt verschiedene Systeme, die die dauerhafte Identifizierung von Ressourcen ermöglichen:

- Universal Resource Name (URN)
- Universal Resource Identifier (URI)
- Persistent Universal Resource Locator (PURL)
- Digital Object Identifier (DOI)
- Handle-System
- Archival Resource Key (ARK).¹⁵¹

Die von Unsworth benannten Primitive “Sampling” und “Selection” beruhen beispielsweise auf der Identifizierung eines digitalen Objekts. Sie stehen hier in engem Zusammenhang. In beiden Fällen sollen nicht nur einzelne Seiten oder Bilder, sondern auch kleinere Einheiten (Ausschnitte, Details) beschrieben und nach bestimmten Kriterien zusammengestellt werden können. Ein Sample ist demnach z. B. eine Liste von Materialien als Ergebnis einer Suche mit einem bestimmten Suchterm. Die Materialien werden entsprechend der Häufigkeit der in ihr vorkommenden Suchterme gerankt. Sampling eignet sich in diesem Sinne auch zur Anzeige der Häufigkeiten in einem Datensatz.¹⁵²

7.5 Discovery

¹⁵⁰ Vgl. Van Horik (2007) Data Curation, S. 118

¹⁵¹ Nach Van Horik (2007) Data Curation, S. 118

¹⁵² Vgl. Unsworth (2000) Scholarly Primitives

Unter den Punkt „Discovery“ fallen Browsing- und Retrievalfunktionen, die speziell unter dem Aspekt verteilter Bestände und der Beziehungen zwischen Objekten und Teilobjekten interessant sind: „Related to collections and searching is scholars' fondness for browsing as a form of information gathering and stimulation, whether in connection with an established project or a new one. Browsing is fundamental to humanities scholars.“¹⁵³ Ein Modell muss komplexe Relationsgefüge abbilden, um alle einem Objekt zugehörigen Elemente und assoziierten Objekte einer Aggregation darstellen zu können.

Die Funktionalität „Discovery“ basiert zum einen auf Relationen und zum anderen auf der Identifizierung digitaler Objekte. Die Identifizierung aller Elemente ist die Voraussetzung für Mehrwert-Dienste wie Nutzungsstatistiken digitaler Objekte und Alert-Dienste, da sie auf der differenzierten Indexierung von Objekten, Teilobjekten und Aggregationen basieren. Mithilfe standardisierter Metadaten können die Objekte und Objektteile beschrieben und typisiert werden und so beispielsweise über das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) zwischen verschiedenen Anbietern ausgetauscht werden. So sind die formale und inhaltliche Klassifikation Grundlage für die Herstellung von Ähnlichkeitsrelationen. Die Anreicherung mit weiteren Metadaten ist abhängig vom jeweiligen Datenprovider.

7.6 Interoperabilität

Ziel der obigen Ausführungen ist die Zusammenführung digitaler Objekte und die Darstellung der Verweise. Das ist in dieser Arbeit unter zwei Aspekten zu sehen:

- der Zusammenstellung von Forschungsdaten nach zufälligen oder ausgewählten Kriterien in einem Forschungsprozess und

¹⁵³ Brockman et al. (2001) Scholarly Work in the Humanities

- der kollaborativen Arbeit, die den Austausch von Forschungsdaten erforderlich macht bzw. die Benutzung von Daten verschiedener Anbieter bzw. in Strukturen wie dem WWW.

Modelle zur strukturellen Beschreibung sollten daher nach Maßgabe der Interoperabilität standardisiert sein. Neben der Explikation von Relationen zwischen digitalen Objekten und ihrer Identifizierung ist daher die Interoperabilität eine weitere funktionale Anforderung.

Interoperabilität auf der Ebene digitaler Objekte dient der Zusammenführung verschiedener Datenstrukturen in eine einheitliche und dadurch dynamische Struktur. Wenn Daten zwischen Systemen ausgetauscht werden sollen, wird zumeist nicht die Strategie verfolgt, die Systeme strukturell interoperabel zu gestalten, sondern es erfolgt eine Anpassung der Beschreibung digitaler Objekte. Ein Austauschformat für digitale Objekte wird auf der konzeptionellen Ebene festgelegt: „These objects are formally specified according to essential characteristics at the conceptual level, and those specifications are articulated in logical models.“¹⁵⁴

Forschungsdaten werden z. B. aus dezentralen Beständen von unterschiedlichen Repositorien in einem Retrievalsystem zusammengeführt.¹⁵⁵ Die Metadaten zu den bereitgestellten digitalen Objekten werden nach der Maßgabe der Interoperabilität mithilfe einer Syntax standardisiert angegeben. Was sich allerdings hinter den einzelnen Metadatenangaben verbirgt, muss in Modellen aufgearbeitet werden. Um Mehrwertdienste (z. B. Alert-Dienste auf Artikel-Ebene) anbieten zu können, muss bei einer dezentralen Datenhaltung auf ein

¹⁵⁴ Thibodeau (2002) Overview of Technological Approaches...

¹⁵⁵ Siehe beispielsweise im Fall von Dokumenten das europäische Repositorien-Netzwerk DRIVER – URL: <http://www.driver-repository.eu/> oder für kulturelle Artefakte die europäische digitale Bibliothek Europeana – URL: <http://www.europeana.eu/portal/>

übergreifendes Modell der digital vorgehaltenen Objekte bzw. Datenstrukturen zurückgegriffen werden.

8 Möglichkeiten der Strukturmodellierung

Dieses Kapitel bietet einen Überblick zu dem Prozess der Modellierung und den Rollen von Modellen in dieser Arbeit, in den digitalen Geisteswissenschaften und in der Informatik. Im Mittelpunkt steht die Typisierung von Relationen zwischen digitalen Objekten. Das Beispiel OAI-ORE für die Modellierung von Webressourcen wird vorgestellt.

8.1 Modellverständnis „Strukturmodell“

Die Menge der sprachlichen Aussagen über alles in der Welt ist mathematisch nicht berechenbar. Maschinell können dagegen nur berechenbare Probleme dargestellt werden. Um reale Vorgänge der Welt maschinenlesbar zu formalisieren, werden Modelle zur Hilfe genommen. Die Modellbildung erfolgt auf Grundlage gemeinsamer Vorstellungen der Modellbildenden über die realen Gegebenheiten, die sogenannte Diskurswelt.¹⁵⁶ Die Strukturmodellierung, wie sie in dieser Arbeit gebraucht wird, beschreibt logische Datenobjekte und ihre Verknüpfungen auf einer konzeptuellen Ebene. Ein Strukturmodell ist demnach ein konzeptuelles Modell, das auf dem gemeinsamen Verständnis und den Vorstellungen basiert, die Wissenschaftler von ihren Forschungsdaten haben und in welchen Zusammenhängen sie gesehen werden. Es abstrahiert Forschungsgegenstände auf Datenobjekte. Durch Relationen zwischen Datenobjekten werden Funktionalitäten der Diskurswelt ausgedrückt, die z. B. wie in dieser Arbeit aus einer hermeneutischen Perspektive auf die Wissenschaftspraxis der Geisteswissenschaften intellektuell erzeugt werden. Die

¹⁵⁶ Die Behauptung, dass Modelle die Realität abbilden, ist Gegenstand vieler Forschungsrichtungen. So gibt es gegenteilige Meinungen aus dem Bereich Philosophie wie beispielsweise von Immanuel Kant, der die These vertritt, dass sich die Realität anhand der Modelle im menschlichen Denken konstituiert. Entsprechende Ausprägungen gibt es auch im Bereich der Sprachwissenschaft, die untersucht, ob Sprache das Denken konstituiert oder umgekehrt das Denken bzw. die Kognition die Sprache beeinflusst.

Konzepte der Funktionalitäten bzw. der Beziehungen von Forschungsgegenständen werden auf Grundlage gemeinsamer Vorstellungen erstellt.

Digitale Objekte können mittels verschiedener Verfahren beschrieben und modelliert werden. Dazu zählen wissensbasierte Verfahren wie inhaltsbeschreibende Metadaten, Indexierungsmittel wie Beschreibungsstandards (Bsp. Text Encoding Initiative (TEI) für Text), Klassifikationen oder ein kontrolliertes Vokabular wie eine Ontologie (Bsp. CIDOC CRM für kulturelle Artefakte). Alle diese Formen zielen auf die semantische Beschreibung, also den Bedeutungsgehalt eines Objekts ab. Sie basiert auf einer Zuweisung von Begriffen zu Objekten bzw. zwischen Begriffen. Die strukturelle Beschreibung identifiziert die einzelnen Objekte, denen die jeweilige Bedeutung zugeschrieben wird und ihre logischen Relationen zu anderen Objekten. Unter der Bezeichnung „Content Model“ werden sowohl die strukturelle als auch die semantische Beschreibung zusammengefasst.

Um den Austausch von Forschungsdaten zu ermöglichen und Verknüpfungen dieser darzustellen, muss eine gemeinsame Definitionsbasis für digitale Objekte vorliegen. Das Konzept der logischen Objekte wurde auf Grundlage des OAIS-Modells erläutert. Die strukturelle Beschreibung mithilfe eines Modells bezieht sich auf logische Objektstrukturen. Durch die Struktur wird im Allgemeinen die logische (An-)Ordnung bzw. der Aufbau einzelner Elemente zu einem Ganzen bestimmt – sie wird in diesem Sinne durch ein Objekt in einem System sowie die Relation zwischen Objekten gekennzeichnet. Diese Relationen konstituieren den Objektraum.

In der Literatur wird die hier beschriebene Strukturmodellierung häufig synonym zu „Objektmodellierung“ verwendet. Mit der Bezeichnung „Objektmodell“ wird allerdings suggeriert, dass sich ein digitales Objekt als Einheit beschreiben lässt. Wie hingegen bereits festgestellt wurde, beschreibt ein Strukturmodell jedoch vielmehr Aggregationen und muss potentielle Verknüpfungen zulassen bzw. für

die Integration von weiteren digitalen Objekten in eine Aggregation offen sein. Daher wurde für diese Arbeit der Terminus „Strukturmodellierung“ gewählt.

8.2 Modelle in der Informatik

Die informatische Modelltheorie¹⁵⁷ besagt, dass ein Modell ein Abbild eines Individuums im Sinne eines Einzelelementes (Atomarität) und seiner Attribute darstellt. Die Attribute können Eigenschaften, Relationen und Operationen sein, die sich auf die atomare Einheit oder ein Attribut beziehen.

Ein Modell weist die folgenden Merkmale auf:

- Es abstrahiert einen Gegenstand (Verkürzungsmerkmal),
- es bildet einen Gegenstand ab (Abbildungsmerkmal)
- und es wird zu einem bestimmten Zweck erstellt (pragmatisches Merkmal).

Da es keine „a priori richtigen oder falschen Modelle“¹⁵⁸ gibt, kann dem Modellschaffenden gemäß des pragmatischen Merkmals eine gewisse Freiheit in der Erstellung zugeschrieben werden. Modelle müssen weiterhin – sofern sie nicht physisch vorliegen – in einer bestimmten Sprache ausgedrückt werden. „Sprache“ wird hier in der Informatik als eine Menge von Zeichen verstanden, die auch Symbole und andere grafische Lösungen beinhalten. Letztlich basieren sie auf natürlicher Sprache, d. h. formale Sprachen sind Modelle natürlicher Sprache. Die Bedeutung der Zeichen natürlicher Sprache bleibt bei der Modellierung erhalten. Dadurch ist es möglich, auch Modelle in semiotischer Hinsicht zu untersuchen. Um die Bedeutung der Zeichen zu verstehen, muss es allerdings Konventionen über ihre Aussagekraft bzw. Gültigkeit geben. Diese Konventionen sind das

¹⁵⁷ Vgl. Glinz (2005) Einführung in die Modelltheorie

¹⁵⁸ Glinz (2005) Einführung in die Modelltheorie, S. 11

bereits zuvor angesprochene Kontextwissen, das dabei hilft, Information zu interpretieren.

Damit Modelle verstanden werden, müssen ebenfalls Konventionen über die Bedeutung der Zeichen des Modells bestehen, d. h. es muss Einigkeit über die Begrifflichkeiten einer Domäne und die Regeln ihrer Anwendung herrschen. Die Regeln und Begriffe bilden einen Kontext, der sich in den Zeichenkonventionen des Modells manifestiert. Sie können mittels des semiotischen Dreiecks¹⁵⁹ abgebildet werden. Das semiotische Dreieck wird zur Hilfe genommen, wenn neben Signifikat („Gezeigtes“) und Signifikant („Zeiger“) noch eine dritte Instanz hinzukommt, die Referenten. Referenten sind die Vertreter des Konzepts („Gezeigtes“, Signifikat) in der Realwelt. Zwischen Signifikant und Referent besteht also nur die indirekte Verbindung über das Konzept (Signifikat).

Es gibt verschiedene Konzepte der Modellierung wie das relationale Modellierungskonzept (z. B. umgesetzt mittels Entity Relationship Model¹⁶⁰) und das objektorientierte Modellierungskonzept (z. B. umgesetzt mittels Unified Modeling Language¹⁶¹). In der objektorientierten Modellierung wird ein abstraktes Objekt erschaffen, das über seine Attribute identifiziert wird, während bei der relationalen Modellierung eine Zeichenkette für ein Objekt steht. Die Attribute können sich ändern und das Objekt trotzdem identifiziert werden. Aus semiotischer Sicht können daher bei objektorientierten Modellen Referent und Signifikant getrennt werden, im relationalen Modell ist der Signifikant selber

¹⁵⁹ Nach Ogden und Richards (1923) *The Meaning of Meaning. A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul

¹⁶⁰ Vgl. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/daten-wissen/Datenmanagement/Daten-/Entity-Relationship-Model->

¹⁶¹ Vgl. <http://www.uml.org/>. Eine Übersicht zu den Relationstypen in UML siehe OAIS Reference Model S. C-1

Referent.¹⁶² Darüber hinaus gibt es das graphenbasierte Modellierungskonzept, das meist zu Wissensrepräsentation benutzt wird. Es basiert auf gerichteten und ausgezeichneten Graphen. Knoten sind Konzepte, deren Beziehungen untereinander durch Kanten modelliert werden. Umgesetzt wird es z. B. mittels RDF oder OWL.¹⁶³

Ergebnis der Modellbildung in der Informatik ist die formale Explikation von Konzepten auf Grundlage eines Modells. Die Konzepte entstehen aus gemeinsamen Vorstellungen über die Diskurswelt. Im vorliegenden Fall werden Relationen zwischen Objekten im pragmatischen Sinne konzeptionalisiert, d. h. funktionale Anforderungen werden umgesetzt. Die Pragmatik wird als der Teil der informatorischen Modellbildung beschrieben, indem der Modellbildende anhand des Originals (Referent im semiotischen Dreieck) ein Konzept (Signifikat) erstellt (siehe auch pragmatisches Merkmal). Die Signifikanten (formale Sprache) werden mithilfe des Modells interpretiert.

8.3 Relationen

Die Auszeichnung digitaler Objekte als Versionen, Auflagen, Formate und Übersetzungen kann meist schon durch standardisiertes Vokabular wie Metadaten (z. B. im Format Dublin Core¹⁶⁴) umgesetzt werden. Die Typisierung der Relation zwischen Objekten in einem übergreifenden Strukturmodell ist darüber hinaus notwendig, wenn z. B. zwei unterschiedliche Objekte Versionen voneinander sind, die nicht gleich klassifiziert werden. Zwischen unterschiedlichen Datenanbietern kann eine Relationsbeschreibung beispielsweise mittels Austauschprotokoll wie dem OAI Protocol for Metadata Harvesting (OAI-PMH) an die einzelnen

¹⁶² Vgl. Margulies (2009) Digitale Daten als Quelle der Geschichtswissenschaft, S. 108

¹⁶³ Vgl. Margulies (2009) Digitale Daten als Quelle der Geschichtswissenschaft., S. 106ff.

¹⁶⁴ Vgl. <http://dublincore.org/>

Datenanbieter weitergegeben werden. Auf Grundlage der Relationsbeschreibung werden die Metadaten zu dem Objekt angereichert.

Spezifizierte Relationen zwischen digitalen Objekten stellen ihren Kontext her. Sie sind Ausdruck der Konzepte wissenschaftlicher Kommunikationsprozesse. Sie sind semantisch motiviert und stellen ein maschinenlesbares Verständnis von Forschungsprozessen her.

Objektrelationen können differenziert werden. So können sie durch das Prinzip der Hierarchie dargestellt werden. Es gibt Vererbungshierarchien (Merkmalsvererbung: Spezialisierung und Generalisierung, Teilmengenbeziehung, Elementbeziehung) und Typhierarchien (Methoden: Ober- und Untertypen, die das Verhalten der modellierten Gegenstände abbilden). Daneben können Relationen über die Zuweisung von Variablen hergestellt werden (relationales Modellierungsprinzip).¹⁶⁵ Weiterhin gibt es argumentative Relationen, thematische Relationen und Ähnlichkeitsrelationen. Ähnlichkeitsrelationen können sich auf formale Merkmale wie Versionen oder Benennungen von Objekten beziehen. Sie können auch inhaltsbasiert sein wie Bild- oder Textähnlichkeit. Die Ähnlichkeitsbestimmung ist abhängig vom Retrievalmodell. Darüber hinaus sind nutzergenerierte Assoziationen bzw. User Generated Relations zu nennen (z. B: referentielle Verknüpfungen wie Zitationen).

Der Unterschied der Strukturmodellierung zum semantischen Netz ist, dass Relationen zwischen digitalen Objekten auf einer logischen Ebene hergestellt werden und nicht zwischen Begriffen.

Für die Typisierung von Relationen in Strukturmodellen kann auf verschiedene Vorarbeiten (Dublin Core (DC), Functional Requirements for Bibliographic

¹⁶⁵ Vgl. Margulies (2009) Digitale Daten als Quelle der Geschichtswissenschaft, S. 81ff.

Records (FRBR)), Fedora) zurückgegriffen werden, die in der ORE-Spezifikation¹⁶⁶ sowie u. a. der Europeana-Spezifikation¹⁶⁷ und einer Publikation des europäischen Repositorien-Netzwerks Driver¹⁶⁸ benannt werden.

8.4 Modellierung im Bereich der digitalen Geisteswissenschaften

Die in dieser Arbeit behandelten funktionalen Anforderungen bilden nur einen Teilbereich der Anforderungen an die Darstellung geisteswissenschaftlicher Forschungsgegenstände im digitalen Raum ab. Willard McCarty widmet sich in einem Standardwerk zum Thema „Humanities Computing“¹⁶⁹ ausführlich der Modellierung von Eigenschaften und Prozessen, die der geisteswissenschaftlichen Forschung und ihren Gegenständen eigen sind. Zu diesen Eigenschaften gehören beispielsweise die Originalität der Forschungsgegenstände (die mithilfe der Datendokumentation nachgewiesen werden kann); die Materialität (durch Nachbildung von Eigenschaften analogen Materials); ihre Authentizität (die beispielsweise in philologischen Wissenschaften über eine Stilanalyse eines Textes oder der Sprache festgestellt wird) oder Authority (womit die „Qualifizierung“ eines Wissenschaftlers gemeint ist, sich zu Forschungsthemen zu äußern). Obgleich die in dieser Masterarbeit vorgestellte Strukturmodellierung einen anderen Fokus hat, ist es unabdingbar das Modellierungsverständnis von Willard McCarty als einem der Hauptakteure des Humanities Computing anzuführen, da seine Arbeiten zu den grundlegenden in den digitalen Geisteswissenschaften zählen.

Ein Modell wird nach McCarty zum Zweck einer Untersuchung eines Gegenstands erschaffen oder dient als Schablone für das Erschaffen von etwas

¹⁶⁶ OAI-ORE (2008) ORE Specifications and User Guides

¹⁶⁷ Dekkers et al. (2009) Europeana Outline Functional Specification

¹⁶⁸ Woutersen-Windhouver (2009) Enhanced Publications: state-of-the-art, S. 47

¹⁶⁹ Vgl. McCarty (2004) Modeling: A Study in Words and Meanings sowie McCarty (2005) Humanities Computing

Neuem. Die Modellierung versteht McCarty in Anlehnung an Clifford Geertz¹⁷⁰ als heuristischen Prozess der Konstruktion und Manipulation eines Modells. Dabei folgt er Geertz, indem das „Modell für Etwas“ (Beispiel: Architektur) vom „Modell von Etwas“ (Beispiel: Sprache) unterschieden wird.¹⁷¹ McCarty widmet sich der Modellierung mit dem Ziel, für neue Ausdrucksformen im Bereich des „Humanities Computing“ ein umfassendes Verständnis zu erbringen. Die „neuen Ausdrucksformen“ verhalten sich dabei ähnlich wie die Einführung des Hypertextes, der neue Formen der Referenzierung ermöglicht hat. Sie müssen nach McCarty durch ein Computersystem explizit gemacht und konsistent dargestellt werden. Durch die Darstellung mithilfe eines Computersystems sieht er die Folgeprobleme der Manipulierbarkeit und Veränderbarkeit digitaler Repräsentation. Bei allen Überlegungen mit dem Ziel der Erstellung adäquater Modelle steht stets die Frage im Mittelpunkt, was sich überhaupt alles maschinenlesbar ausdrücken lässt. McCarty verdeutlicht dieses Anliegen im nachfolgenden Beispiel:

„Take, for example, knowledge one might have of a particular thematic concentration in a deeply familiar work of literature. In modeling one begins by privileging this knowledge, however wrong it might later turn out to be, then building a computational representation of it, e.g., by specifying a structured vocabulary of word-forms in a text-analysis tool. In the initial stages of use, this model would be almost certain to reveal trivial errors of omission and commission. [...]”¹⁷²

In Anlehnung an Brian Cantwell Smith¹⁷³ formuliert McCarty, dass Modelle die Grundlage der elektronischen Datenverarbeitung (Computing) sind:

¹⁷⁰ Vgl. Geertz, Clifford (1973) *The Interpretation of Cultures. Selected Essays*. Basic Books Classics

¹⁷¹ Vgl. McCarty (2004) *Modeling: A Study in Words and Meanings*, S. 255

¹⁷² McCarty (2004) *Modeling: A Study in Words and Meanings*, S. 255

¹⁷³ Vgl. Smith, Brian Cantwell (1995/1985) *Limits of Correctness in Computers*. In: Johnson, Deborah & Nissenbaum, Helen [Hrsg.] (1995/1985) *Computers, Ethics and Social Values*, S456-69, zitiert bei McCarty (2004) *Modeling: A Study in Words and Meanings*, S. 255

„[...] to do anything useful at all a computer must have a model of something, real or imaginary, in software. But in the context of computing, models *per se* are not the point. What distinguishes computers from other kinds of machines, Smith notes, is that “they run by *manipulating* representations and representations are always formulated in terms of models.”¹⁷⁴

McCarty betrachtet die Aktivitäten der digitalen Geisteswissenschaften, um sie zu konzeptionalisieren und in ein Modell zu überführen. Das Modell kann stets einem Stadium der Wissensgewinnung über einen Gegenstand entsprechen, stellt aber keine fixen Wissensstrukturen dar. Modelle werden an ihrer Konsistenz und Explizitheit gemessen. Die Frage hierbei ist, wie explizit man Gegenstände in Modellen überhaupt machen kann. Wichtig am Modellverständnis McCartys ist, dass auch die Strukturmodelle, die in dieser Arbeit behandelt werden, einer wissenschaftlichen Praxis gerecht werden müssen, indem Forschungsgegenstände mitsamt ihrem Kontext maschinenlesbar dargestellt werden. Hierbei gilt es, wie die Benennung exemplarischer funktionaler Anforderungen gezeigt hat, insbesondere Relationen zwischen digitalen Objekten zu explizieren und Objekte zu identifizieren.

8.5 OAI „Object Reuse and Exchange“

Die Initiative ORE will Standards und Protokolle zum Datenaustausch zwischen Anbietern von Informationsressourcen unter der Maßgabe der Interoperabilität bereitstellen, um ihre nachhaltige Nutzbarkeit zu gewährleisten. Im Fokus des Standards „Object Reuse and Exchange“ (ORE) der „Open Archives Initiative“ (OAI) steht der Austausch von Webressourcen zwischen Repositorien. Die ORE-Spezifikation wurde 2008 veröffentlicht und liefert ergänzende Angaben zum Standard „OAI Protocol for Metadata Harvesting“ (OAI-PMH). Sie orientiert sich an den Web-Standards des World Wide Web Consortiums (W3C)¹⁷⁵

¹⁷⁴ McCarty (2004) Modeling: A Study in Words and Meanings, S. 257

¹⁷⁵ Vgl. <http://www.w3.org/>

untermöglicht die Formalisierung auf verschiedene Weise (Atom, RDF/XML, RDFa).¹⁷⁶

ORE ist ausgehend von der Fragestellung entwickelt worden, woher ein Datenanbieter die notwendigen Informationen über Struktur und Kontext eines digitalen Informationsobjekts beziehen soll. Ziel ist, digitale Informationsobjekte mitsamt seiner Komponenten zu identifizieren¹⁷⁷: „As the objects get more complex, things get worse.“¹⁷⁸ Mit diesen Worten umschreibt Herbert Van de Sompel, einer der Entwickler von ORE, die Ausgangsproblematik der Darstellung komplexer digitaler Objekte. Ein Problem stellt beispielsweise deren konsistente Verlinkung dar. So steht in den Geisteswissenschaften selten eine Informationsressource für sich allein. Soll eine funktionale Anforderung wie „Discovery“ ermöglicht werden, so müssen alle Komponenten einer Aggregation gefunden werden können. Eine Suchmaschine kann ohne entsprechende Informationen die Komponenten eines digitalen Informationsobjekts nicht ohne weiteres zuordnen. Sind mehrere Versionen eines Objekts vorhanden, die zur selben Aggregation gehören, muss es möglich sein, darzustellen, auf welche Komponente (bzw. Version) verwiesen wird bzw. auf eine bestimmte Komponente zu verweisen.

ORE bezeichnet Webressourcen als Aggregationen digitaler Informationsobjekte („compound digital objects“). Der Begriff der „Aggregation“ ist der Versuch, zusammenhängende Informationsressourcen in einem logischen Ganzen darzustellen, das für Webanwendungen erkennbar ist. Eine Aggregation sowie

¹⁷⁶ Vgl. „Atom“ – URL <http://www.atompub.org/>; „RDF“ – URL <http://www.w3.org/RDF/>; „RDFa“ –URL: <http://rdfa.info/>

¹⁷⁷ Neben ORE gibt es weitere Standards wie den „Metadata Encoding & Transmission Standard“ (METS – Vgl.: <http://www.loc.gov/standards/mets/>) als Containerformat zur Kapselung digitaler Objekte und ihrer Metadaten oder die „Digital Item Declaration Language“ (DIDL – Vgl. <http://xml.coverpages.org/mpeg21-didl.html>) als Beschreibungssprache der „Moving Pictures Expert Group“ (MPEG) für strukturierte digitale Objekte, die Multimedia-Elemente enthalten.

jede ihrer Komponenten muss als eigene Ressource identifiziert werden können (mithilfe einer ID wie einem Persistent Identifier). Die einzelnen Komponenten können unterschieden werden

- nach „Genre“ oder Dokumenttyp („semantic type“),
- nach Medientypen und ihren Formaten (Bsp. Text im PDF-Format oder eine Audiodatei im mp3-Format),
- die internen Strukturen (z. B. ist eine Komponente eine weitere Ansicht einer anderen Komponente („multiple views“)),
- die externen Strukturen, die wiederum unterteilt werden können
 - in Referenzierungsbeziehungen wie Zitation,
 - Verweis auf weitere Versionen einer Komponente,
- und in ihren Metadaten.

Die Darstellung der Aggregationen erfolgt durch die Implementation einer ‚Resource Map‘. Eine ORE Resource Map beinhaltet die ORE-Repräsentation einer ORE-Ressource mit allen zugehörigen Relationen. Diese identifiziert jede Aggregation sowie die dazugehörigen Ressourcen eindeutig mit Hilfe einer URI. Die Aggregation und alle Ressourcen werden durch Metadaten beschrieben. ORE drückt interne Verknüpfungen von logisch zusammenhängenden Aggregaten sowie externe Verknüpfungen von logisch zusammenhängenden Objekten durch sogenannte „typed relations“ aus.

¹⁷⁸ Van de Sompel et al. (2007) Open Archives Initiative Object Re-Use & Exchange

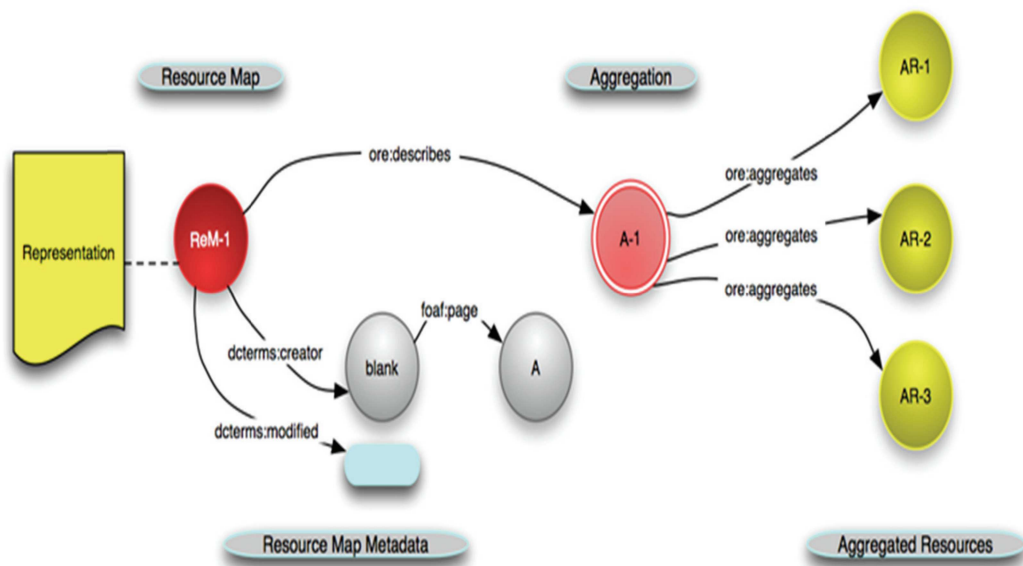


Abbildung 6: ORE Resource Map¹⁷⁹

Typed Relations können mithilfe von RDF-Tripeln oder RDF-Graphen formalisiert werden. Sie basieren auf einem hierarchischen Prinzip. Eine Spezifizierung der Relationen ist in Planung.¹⁸⁰

ORE bezieht sich mit Webressourcen auf einen Ausschnitt des Spektrums digitaler Forschungsdaten für die Geisteswissenschaften. Es setzt die Grundidee der (hierarchischen) Explikation von Relationen zwischen digitalen Objekten sowie deren Identifizierung um und wird daher als passender Ansatzpunkt der Strukturmodellierung gesehen. Die Schlussfolgerung lautet nicht, dass ORE ein paradigmatisches Beispiel für die strukturelle Beschreibung digitaler Forschungsdaten im Bereich der digitalen Geisteswissenschaften ist.

¹⁷⁹ http://www.openarchives.org/ore/1.0/primer-images/ADM_basics.png

¹⁸⁰ Dafür werden beispielsweise die Relationstypen des ISO-Standards „CIDOC Conceptual Reference Model“ (CRM – Vgl. <http://cidoc.ics.forth.gr/index.html>), der Functional Requirements on Bibliographic Records“ (Vgl. <http://www.ifla.org/en/publications/functional-requirements-for->

ORE unterstützt die maschinenlesbare Darstellung und Auswertung der Relationen zwischen digitalen, komplexen Objekten, das Auffinden dieser durch Webanwendungen sowie den Datenaustausch zwischen verschiedenen Anbietern: „The objective of the OAI-ORE Initiative is to develop an interoperability layer across cooperating digital repositories, registries and services for the reuse and exchange of compound digital objects, based on the Web architecture [...]”¹⁸¹

Die Modellierung auf Grundlage von ORE vereinfacht das Publizieren von Informationsobjekten, die Bereitstellung von Mehrwertservices wie Discovery Services und Linking und ihre Prozessierung in anderen Systemen. Über den Austausch von Daten auf Grundlage von ORE kann die Darstellung dieser in Webanwendungen wie Suchmaschinen, kollaborativen Plattformen wie Virtuellen Forschungsumgebungen und Workflow-Tools gelingen.

Sowohl das europäische Repository-Netzwerk DRIVER als auch das Projekt Europeana haben sich dazu entschlossen, die OAI-ORE-Spezifikation als Grundlage zur Modellierung ihrer Daten zu verwenden. Beide Projekte und auch die Akteure der ORE-Initiative sehen die Notwendigkeit, eine Spezifikation für Linktypen zu erstellen.

Ein bereits angesprochenes Problem ist die Darstellung der Grenzen komplexer digitaler Objekte. Auch ORE kann die Grenzen einer logischen Einheit nicht ausdrücken bzw. maschinenlesbar formalisieren.

bibliographic-records) oder des Dublin Core Metadata Initiative Abstract Model (DCMI – Vgl. <http://dublincore.org/documents/abstract-model/>) in Erwägung gezogen.

¹⁸¹ Cheung et al. (2008) SCOPE, S. 6

8.6 Zusammenfassung

Die im Titel der Masterarbeit formulierten „Möglichkeiten“ der Strukturmodellierung beziehen sich nicht auf einzelne Beispiele wie ORE. Sie beziehen sich vielmehr auf die technischen Möglichkeiten, die sich mithilfe der strukturellen Modellierung ergeben, sowie die Potentiale, die daraus für die geisteswissenschaftliche Forschung entstehen.

Modelle basieren auf gemeinsamen Vorstellungen bzw. Vereinbarungen über einen Gegenstand. Es können mehrere Folgerungen gewonnen werden:

- Aus syntaktischer Sicht kann alles, was aufgrund eines gemeinsamen Datenformats als logisches Objekt betrachtet wird, als digitales Objekt betrachtet werden. Im konkreten Fall wurde als Arbeitsdefinition das OAIS-Modell herangezogen. Es trifft hingegen keine Aussage dazu, wann aus einem digitalen Objekt ein bedeutungstragendes Objekt wird.
- Aus einer pragmatischen Sicht ist ein digitales Objekt dann digitales Objekt, wenn es nach einem Modell als solches beschrieben werden kann.
- Aus einer semantischen Sichtweise basieren digitale Objekte als Forschungsdaten auf der Annahme, dass alles Forschungsobjekt sein kann, was Gegenstand einer wissenschaftlichen Untersuchung ist. Forschungsdaten sind auf einer konzeptuellen Ebene zu sehen.

Eine Zusammenfassung lässt folgenden Schluss zu: Aussagen zu konzeptuellen Ebene sind semantisch motiviert. Als Beispiel dafür dienen die Relationen, die typisiert werden. Die Typisierung erfolgt auf Basis funktionaler Anforderungen, die aus der hermeneutisch geprägten Wissenschaftspraxis erwachsen. Relationen werden beschrieben. In einem Modell werden den Relationen Bedeutungen aus der Diskurswelt in natürlicher Sprache zugeschrieben. In der Implementierung wird sie wieder auf einer syntaktischen Ebene formalisiert (formale Sprache). Somit wird Wissen aus der Diskurswelt angewendet, syntaktisch kodiert, um dann

von einem Empfänger entschlüsselt zu werden. Es ist damit Wissen aus Sicht der Shannonschen Informationstheorie.¹⁸²

Forschungsdaten sind aus Sicht der Semiotik Bedeutungsträger. Durch die Bezugnahme aufeinander wird zur Bedeutungsfindung beigetragen. Die Relationen, die die Bezugnahme ausdrücken, sind selbst semantisch motiviert und daher Bedeutungsträger. In einer Erweiterung nach dem semiotischen Dreieck kann ein Strukturmodell als Konzept (Signifikat) bezeichnet werden. Die bezeichneten Objekte, d. h. Forschungsdaten sind Bedeutungen des Konzepts und damit Referenten. Die Beschreibungssprachen auf einer syntaktischen Ebene sind Signifikanten.

Die beiden zentralen Aspekte bei der Modellierung von digitalen Objekten sind sicher Komplexität und Kontext. Beide stehen in enger Wechselwirkung: der Komplexität kann durch Kontextualisierung begegnet werden und andererseits führt sie selbst wieder zu neuer Komplexität.

Die Kontextualisierung der digitalen Informationsobjekte muss auf einer konzeptuellen Ebene modelliert und auf einer syntaktischen Ebene umgesetzt werden. Funktionen wie der Vergleich von Forschungsgegenständen, die Neuzusammenstellung und „das Entdecken“ (Browsing und Retrieval) von neuen Forschungsgegenständen sollten unterstützt werden. Diese Funktionen basieren auf einer Identifizierung von Objekten und Teilobjekten bzw. Komponenten einer Aggregation. Weiterhin müssen zur Kontextualisierung Relationen zwischen digitalen Objekten dargestellt werden können, um beispielsweise abzubilden, dass ein digitales Objekt eine Version, eine Annotation oder eine Transkription eines anderen digitalen Objekts ist.

¹⁸² Vgl. Voß (2009) Zur Neubestimmung des Dokuments im rein Digitalen

Die strukturelle Interoperabilität auf Objektebene auf Grundlage von OAI-ORE wird in diesem Zusammenhang als Beispiel verwendet, weil

- der Datenaustausch via WWW
- und die Verknüpfungsmöglichkeiten zwischen Webressourcen

„Vorbildcharakter“ für den Umgang mit digitalen Forschungsdaten der Geisteswissenschaften hat. ORE bietet hier einen Ansatz – der freilich noch zu modifizieren und auszubauen ist – der auf dem OAI-PMH-Protokoll basiert und die Interoperabilität als Grundvoraussetzung für Nachnutzbarkeit von Forschungsdaten zur Maßgabe hat.

ORE ist zudem in der Lage, verschiedene Stadien der wissenschaftlichen Wertschöpfung von Daten einzubeziehen. So ist in der Beschreibung zum DRIVER-Projekt nachzulesen, dass ORE generisch ist, da es ergänzende Relationen bzw. die Einbindung digitaler Objekte zulässt.¹⁸³ Perspektivisch ist ein Modell wie ORE auf die Zusammenführung heterogener Auszeichnungsschemen und zwischen heterogenen Datensystemen angelegt.

¹⁸³ Vgl. <http://www.driver-repository.eu/Enhanced-Publications.html>

9 Abschlussbetrachtung: Potentiale

Die Masterarbeit hat sich mit den Möglichkeiten der Verknüpfung digitaler Forschungsdaten im Forschungsprozess der hermeneutisch geprägten Geisteswissenschaften auseinandergesetzt. Durch die gegenseitige Bezugnahme der Forschungsgegenstände sowie die Einbettung in neue inhaltliche Kontexte entsteht ein komplexes Beziehungsgefüge zwischen digitalen Objekten. Aus technischer Sicht wird der Kontext von digitalen Objekten modelliert, um sie in ihrer Komplexität maschinenlesbar darzustellen. Die Modellierung basiert auf typisierten Relationen, die Beziehungen zwischen Forschungsdaten herstellen. Ziel ist die Nachnutzung der Daten, die Nachvollziehbarkeit des Forschungsprozesses und der Transfer verteilter Datenbestände in digitalen vernetzten Umgebungen.

Die Ausführungen zeigen, dass durch die formal explizierte und mediale Zusammenführung von digitalen Forschungsdaten sowie die Zusammenführung der Gegenstände und Instrumente neue Wege des wissenschaftlichen Diskurses ermöglicht werden. Die Bezugnahme selbst ist Instrument der Forschung. Der wissenschaftliche Diskurs und die Bezugnahme können mit digitalen Forschungsdaten präziser dokumentiert und durch Wissenschaftler in anderer Form als der traditionellen nachvollzogen werden. Mithilfe der Strukturmodellierung kann den funktionalen Anforderungen an die Bereitstellung digitaler Forschungsdaten, die aus der Wissenschaftspraxis entstehen, Rechnung getragen werden.

Die strukturelle Verknüpfung digitaler Objekte, so das Ergebnis und gleichzeitige These für nachfolgende Arbeiten, schafft Wissen, da die Interpretation von Zeichen auf inhaltlich motivierten Relationen basiert, die damit Bedeutung tragen. Struktureller Kontext schafft demnach zugleich Bedeutungskontext. Eine Diskussion dieser These ist wünschenswert, da sie mit den oben genannten Aspekten zum wissenschaftlichen Diskurs zu der Behauptung führt, dass neue Wege der Wissensproduktion beschritten werden. Gleichzeitig gewinnen

Wissenschaftler ein neues Verständnis vom Prozess der Wissensproduktion. Die Möglichkeiten der Verknüpfungen erweitern den strukturellen wie auch den Bedeutungskontext, so dass sich das Spektrum der Bearbeitungs- und Untersuchungsmöglichkeiten dadurch ebenfalls vergrößert – dies gilt im Besonderen für Virtuelle Forschungsumgebungen, deren Ziel die Bereitstellung von Werkzeugen ist. Hier liegt das Potential für neue Forschungsansätze in den digitalen Geisteswissenschaften.

Perspektivisch wird zum einen auf die Ziele des Semantic Web verwiesen, das mithilfe fester Syntax und Semantik Bedeutungszusammenhänge zwischen Begriffen und Objekten herstellt. Darüber hinaus ist die Bedeutung der Strukturmodellierung unbestritten, findet sich in ihr doch der Ansatz des von Tim Berners-Lee geprägten „Data Web“ bzw. „Linked Data“ wieder.¹⁸⁴ Die Erkenntnisse dieser Arbeit decken sich mit diesem Konzept des „Webs der Zukunft“, in dessen Mittelpunkt die mithilfe von RDF formalisierten Verweise zwischen strukturierten Daten stehen. Die Verweise sind nicht lokal orientiert, sondern basieren auf der Identifizierung von digitalen Objekten. Auf der Basis typisierter Relationen wird Bedeutung hergestellt. Oder mit den Worten von Tim Berners Lee:

„[...] data is relationships [...] the more things you have to connect, the more powerful is raw data [...]!“¹⁸⁵

Das Data Web beinhaltet nicht nur wissenschaftliche Daten, wie sie in dieser Arbeit von Interesse waren, sondern verschiedenste Datenformen. Eine weitgreifende Behauptung ist, dass alle symbolischen Artefakte zum Forschungsgegenstand der Geisteswissenschaften werden können und damit auch

¹⁸⁴ Vgl. <http://www.w3.org/DesignIssues/LinkedData.html>

¹⁸⁵ Berners-Lee, Tim (2009) Tim Berners-Lee on the next Web (TED-Talk) – URL: http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html Tim

alles was im „Data Web“ auf einer konzeptionellen Ebene dargestellt wird. Das nominalistische Zeichenverständnis der Informatik, das Verweise zwischen Begriffen und Objekten herstellt, muss erweitert werden: Objekte sind Bedeutungsträger und verweisen auf Objekte als Bedeutungsträger. Es entsteht ein komplexes Beziehungsgefüge, in dem der Begriff des Zeichens mehrdimensional betrachtet werden muss.

Das Konzept der digitalen Geisteswissenschaften sowie des e-Research besitzen großes Potential über die reine Bereitstellung von Daten und Infrastrukturen hinaus. Die zu Beginn erwähnte, von NSF und JISC formulierte, Cyberscholarship kann in beiden Ausprägungen unterstützt werden: Es entstehen neue Formen der Wissenschaftspraxis („new forms of scholarship“), in denen die digitalen Objekte selbst zu Infrastrukturelementen werden („content as infrastructure“) und die Infrastrukturelemente wiederum zu möglichen Ressourcen.

Das eingangs genannte Attribut für digitale Forschungsdaten in den Geisteswissenschaften kann bestätigt werden: Sie sind komplex. Daneben wächst ihre Zahl in quantitativer Hinsicht mit jeder Verknüpfung, da sich potentielle neue Wege der Bedeutungsfindung ergeben. Der Austausch von Forschungsdaten bzw. des Wissens über ihre Zusammenhänge ist ein Ansatzpunkt der für die digitalen Geisteswissenschaften erwünschten Kooperation. Die Möglichkeit des Daten- und Wissensaustausches bietet zugleich ideale Ausgangsbedingungen für interdisziplinäre Forschungsprojekte.

Peter Schirnbacher bezieht sich in seinen Ausführungen zu einer neuen Kultur des elektronischen Publizierens¹⁸⁶ auf die durch Informations- und Kommunikationstechnologien veränderten Prozesse in der

¹⁸⁶ Vgl. <http://edoc.hu-berlin.de/cmsj/27/schirnbacher-peter-19/PDF/schirnbacher.pdf>

Wissenschaftskommunikation. Im Mittelpunkt steht hier das wissenschaftliche elektronische Publizieren und die „Kultur“ meint die dabei geltenden Regeln und Verhaltensweisen. Die Möglichkeiten der Einbindung von digitalen Forschungsdaten sowohl in die wissenschaftlichen Arbeitsprozesse, den Diskurs und die wissenschaftlichen Publikationen selbst legen die Vermutung nahe, dass sich die Regeln und Verhaltensweisen weiter verändern werden. Eine Untersuchung der sich daraus ergebenden Potentiale in Bezug auf das wissenschaftliche elektronische Publizieren bildet einen Ansatzpunkt für zukünftige Forschungsarbeiten. Die Rolle von Open Access zu wissenschaftlichen Publikationen, die einen Teil der Forschungsdaten für die Geisteswissenschaften bilden, sowie die Bemühungen um den freien Zugang zu wissenschaftlichen Forschungsdaten unter dem Schlagwort „Open Data“ könnten in eine solche Untersuchung einbezogen werden.

Der Fokus des aktuellen Fachdiskurses zum Thema Forschungsdaten in Deutschland liegt nicht allein auf den Möglichkeiten des (freien) Zugangs zu digitalen Forschungsdaten, jedoch nehmen sie großen Raum ein. Die qualitativ hochwertige Aufbereitung der Forschungsdaten in Zusammenarbeit mit den Fachdisziplinen sowie die Verbreitung und Kommunikation der Möglichkeiten, die digitale vernetzte Umgebungen hinsichtlich neuer Wege der Forschung bieten, werden als ebenso wichtig angesehen. Die Magisterarbeit versteht sich in diesem Sinne als Problembeschreibung.

Konkrete neue Forschungsfragen sind nur durch die Wissenschaftler selbst zu entwickeln, die technischen Lösungen werden von der Informatik umgesetzt. Der Bereich der digitalen Geisteswissenschaften besitzt durch die Verschmelzung dreier wissenschaftlicher Domänen ein großes Entwicklungspotential: Das methodische Wissen der Geisteswissenschaften wird mit der praktischen Kompetenz der Informatik und den Erkenntnissen der Bibliotheks- und Informationswissenschaft in der Funktion des Mediators kombiniert.

10 Literaturverzeichnis

10.1 Ausgewertete Literatur- und Internetquellen

Alle Internetquellen wurden zuletzt am 02.08.2009 aufgerufen.

ACLS Commission on Cyberinfrastructure [Hrsg.] (2006) Our Cultural Commonwealth. The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences – URL: http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf

Alexander von Humboldt-Stiftung [Hrsg.] (2008) Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen (Diskussionspapier der Alexander von Humboldt-Stiftung 12) – URL: http://www.humboldt-foundation.de/pls/web/wt_show.text_page?p_text_id=7839

Allianz der Wissenschaftsorganisationen [Hrsg.] (2008) Schwerpunktinitiative „Digitale Information“ der Allianz-Partnerorganisationen. Beschluss der Allianz der Wissenschaftsorganisationen vom 11. Juni 2008 – URL: http://www.hrk.de/de/download/dateien/080603_Beschlussvorlage_gekuerzt_Veroeffentlichung.pdf

Allianz-Initiative “Digitale Information” [Hrsg.] (2008) Schwerpunktinitiative “Digitale Information” der Allianz-Partnerorganisationen (Pressemitteilung) – URL: http://www.dfg.de/aktuelles_presse/das_neueste/download/pm_allianz_digitale_information_details_080612.pdf

Arms, William Y.; Larsen, Ronald L. (2007) The Future of Scholarly Communication. Building the Infrastructure for Cyberscholarship. Report of a workshop held in Phoenix, Arizona April 17-19 (NSF / JISC) – URL: <http://www.sis.pitt.edu/~repwkshop/SIS-NSFReport2.pdf>

Arnsward, Ulrich; Nida-Rümelin, Julian; Arnsward, Wilfried [Hrsg.] (2005) Die Zukunft der Geisteswissenschaften. [Wilfried Arnsward zum 70. Geburtstag]. Heidelberg: Manutius-Verlag (Schriften des European Institute for International Affairs, Heidelberg)

Beiner, Marcus (2009) Humanities. Was Geisteswissenschaft macht. Und was sie ausmacht. Berlin: Berlin University Press

Blanke, Tobias (2006) E-Science in the Arts+Humanities (Präsentation im Rahmen der Tagung “e-Science und Grid-Aktivitäten in Göttingen” am 23.11.06) – URL: http://www.textgrid.de/fileadmin/TextGrid/konferenzen_vortraege/goegrid06/GoeGridSeminar06-AHeSSC.pdf

- Blanke, Tobias (2009) Perspektiven der Technikentwicklung für Geistes- und Kulturwissenschaften (Präsentation im Rahmen der Tagung „HyperImage-Konferenz vom 26. bis 27. Februar 2009“ am 27.02.09) – URL: http://hyperimage.hu-berlin.de/files/Perspektiven_Blanke.pdf
- Brandt, Reinhardt (2005) Zustand und Zukunft der Geisteswissenschaften. In: Arnswald, Ulrich; Nida-Rümelin, Julian; Arnswald, Wilfried [Hrsg.] (2005): Die Zukunft der Geisteswissenschaften. [Wilfried Arnswald zum 70. Geburtstag]. Heidelberg: Manutius-Verl. (Schriften des European Institute for International Affairs, Heidelberg), S. 29-32
- Brase, Jan (2009) Der Digital Object Identifier (DOI). In: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine et al. [Hrsg.] (2009) nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0, Kapitel 9.4.2., S. 57-64 – URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_335.pdf
- Brase, Jan; Haak, Susanne (2008) Management von Primärdaten. Neue Kooperation der Technischen Informationsbibliothek mit dem Verlag Georg Thieme. In: Verfügbarkeit von Information (Proceedings 60. Jahrestagung der Deutschen Gesellschaft für Informationswissenschaft und -praxis (DGI)), S. 83-86
- Brockman, William S.; Neumann, Laura; Palmer, Carole L. et al. [Hrsg.] (2001) Scholarly Work in the Humanities and the Evolving Information Environment. Washington D.C.: Digital Library Federation, Council on Library and Information Resources – URL: <http://www.clir.org/pubs/reports/pub104/pub104.pdf>
- Burkhardt, Hans (2008) Geisteswissenschaften. Geist schafft Wissen. München: Hans-Seidel-Stiftung e.V. Akad. für Politik und Zeitgeschehen (Aktuelle Analysen / Hanns-Seidel-Stiftung, Akademie für Politik und Zeitgeschehen, 47)
- Buckland, Michael (1997) What is a “document”? – URL: <http://people.ischool.berkeley.edu/~buckland/whatdoc.html> (Preprint)
- Consultative Committee for Space Data Systems (CCSDS) (2002) Reference Model for an Open Archival Information System (OAIS) – URL: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Cheung, Kwok; Lashtabeg, Anna; Drennan, John (2008) SCOPE. A Scientific Compound Object Publishing and Editing System. In: The International Journal of Digital Curation 2 (3), S. 4-18 – URL: <http://www.ijdc.net/index.php/ijdc/article/view/84>
- Choudhury, Sayeed (2007) The Relationship between Data and Scholarly Communication. – URL: <http://www.sis.pitt.edu/~repwkshop/papers/choudhury.doc>

- Crane, Gregory (2008) Repositories, Cyberinfrastructure and the Humanities. In: EDUCAUSE Review Magazine 43 (6) – URL: <http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume43/RepositoriesCyberinfrastructure/163269>
- Day, Michael; Guercio, Maria; Herrmann, Volker et al.(2006) Preservable Digital Objects as Building Blocks of Digital Libraries (PDO) – URL: <http://old.hki.uni-koeln.de/people/herrmann/forschung/PDO06-04.pdf>
- Dempsey, Lorcan (2006) The (Digital) Library Environment. Ten Years After. In: Ariadne 46 – URL: <http://www.ariadne.ac.uk/issue46/dempsey/>
- DFG (1998) Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Denkschrift – URL: http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- DFG (o. J.) Forschungsprimärdaten – URL: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/digitale_information/primaerdaten/index.html
- DINI e. V. (2009) Positionspapier Forschungsdaten (Arbeitsgruppe „Elektronisches Publizieren“) – URL: <http://edoc.hu-berlin.de/series/dini-schriften/2009-10/PDF/10.pdf>
- Dekkers, Makx; Gradmann, Stefan; Meghini, Carlo (2009) Europeana Outline Functional Specification. For development of an operational European Digital Library (EDLnet Foundation D2.5) – URL: http://dev.europeana.eu/public_documents/EDLnet%20D2.5_Outline_Functional_Specifications20090301_version%201.7_consWithoutHistory_lossless.pdf
- European Science Foundation - Standing Committee for the Humanities (2007) Position Paper 2007 . – URL: http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/research_areas/HUM/Frontpage/ESF_SCHPosition_24P_Final%20version_V11.pdf&t=1250180579&hash=bd844a948b84cf116b144af32a35b44f
- F&E Rahmenprogramm 2005-2009 (2004) e-Science in Deutschland. F&E Rahmenprogramm 2005-2009 – URL <http://grid.desy.de/d-grid/RahmenprogrammEndfassung.pdf>
- Fraser, Michael (2005) Virtual Research Environments. Overview and Activity. In: Ariadne 44 – URL: <http://www.ariadne.ac.uk/issue44/fraser/>
- Funk, Stefan E. (2009) Digitale Objekte und Formate. In: Neuroth, Heike; Oßwald, Achim; Scheffel et al. [Hrsg.] (2009) nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0, S. 3-8 – URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_352.pdf

- Gersmann, Gudrun, Mruck, Katja [Hrsg.] (2004) Elektronisches Publizieren & Open Access. Historical Social Research (HSR) Sonderheft. 1 (29) – URL: <http://hsr-trans.zhsf.uni-koeln.de/hsrretro/docs/artikel/artikelliste.php>
- Gethmann, Carl Friedrich; Langewiesche, Dieter et al. (2005) Manifest Geisteswissenschaft – URL: http://www.uni-tuebingen.de/uni/f07/download/manifest_geisteswissenschaften.pdf
- Gold, Anne (2007) Cyberinfrastructure, Data, and Libraries, Part 1. A Cyberinfrastructure Primer for Librarians. In: D-Lib Magazine 9/10 (13) – URL: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>
- Glinz, Martin (2005) Einführung in die Modelltheorie. ("Informatik II: Modellierung"). Zürich – URL: http://www.ifi.uzh.ch/rerg/fileadmin/downloads/teaching/courses/infII_ss06/inf_I_I_kapitel_02.pdf
- Gradmann, Stefan; Meister, Jan Christoph (2008) Digital document and interpretation. Re-thinking "text" and scholarship in electronic settings. In: Poiesis Prax, 5, S. 139-153 – URL: <http://www.springerlink.com/content/g370807768tx2027/fulltext.pdf>
- Gradmann, Stefan (2008) Modeling Complex Digital Information Objects. Minutes of a working group bringing together the Europeana, DRIVER, OAI-ORE, OAPen and EERQI projects (Draft version)
- Gradmann, Stefan (2007) Some thoughts on the importance of open source and open access for emerging digital scholarship. In: Hrachovec, Herbert; Pichler, Alois [Hrsg.] (2007) Philosophy of the Information Society. Proceedings of the 30. International Ludwig Wittgenstein Symposium Kirchberg am Wechsel, Austria, Volume 2, S. 275-278
- Gradmann, Stefan (2007) Verbreitung vs. Verwertung. Anmerkungen zu Open Access, zum Warencharakter wissenschaftlicher Informationen und zur Zukunft des elektronischen Publizierens. In: Havemann, Frank, Parthey, Heinrich, Umstätter, Walther [Hrsg.] (2007) Integrität wissenschaftlicher Publikationen in der Digitalen Bibliothek (Wissenschaftsforschung Jahrbuch 2007), Gesellschaft für Wissenschaftsforschung e.V. (GeWiF Berlin), S. 93-106 – URL: http://www.wissenschaftsforschung.de/JB07_93-106.pdf
- Gradmann, Stefan (2005) Gibt es 'Digitale Bibliotheken'? Wird es sie jemals geben? Zu den Grenzen einer allzu populären Metapher. In: Segeberg, Harro; Winko, Simone [Hrsg.] (2005) Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm Fink Verlag, S. 295-313 – URL: http://www1.uni-hamburg.de/DigiLit/gradmann/digitale_bibliotheken.html

- Gradmann, Stefan (2004) Vom Verfertigen der Gedanken im digitalen Diskurs. Versuch einer wechselseitigen Bestimmung hermeneutisch und empirizistischer Positionen. In: Gersmann, Gudrun, Mruck, Katja [Hrsg.] (2004) Elektronisches Publizieren & Open Access. Historical Social Research (HSR) Sonderheft. 1 (29), S. 56-63 – URL: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-30861>
- Greenstein, Daniel (2003) Information Technology and Humanities Scholarship. Humanities scholars are moving toward a radical change in how they conduct research. In: Educause Quarterly, 1, S. 60-61 – URL: <http://net.educause.edu/ir/library/pdf/eqm03110.pdf>
- Hammwöhner, Rainer (2004) Hypertext. In: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar [Hrsg.] (2004) Grundlagen der praktischen Information und Dokumentation (5. Auflage), K.G. Saur: München, S. 419-427
- Hiebler, Heinz (2005) Von der Medienkulturgeschichte digitaler Codierungen zu einem Analysemodell ‚digitaler Literatur‘. In: Segeberg, Harro; Winko, Simone [Hrsg.] (2005) Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm Fink Verlag, S. 85-110 – URL: http://www1.uni-hamburg.de/DigiLit/hiebler/digit_medienkulturgeschichte.html
- Hockey, Susan (2004) The History of Humanities Computing. In: Schreibman, Susan; Siemens, Ray; Unsworth John [Hrsg.] (2004) A companion to digital humanities. Malden, Mass.: Blackwell Publ. (Blackwell companions to literature and culture, 26), S. 3-19 – URL: <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=default>
- Huth, Karsten (2009) Einführung. In: Neuroth, Heike; Oßwald, Achim; Scheffel et al. [Hrsg.] (2009) nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Kapitel 9.1, S. 1-18 – URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_340.pdf
- Jahn, Najko (2009) Wenn das Medium zum Inhalt wird. Bericht über den Abschluss-Workshop des DFG-Projekts „Konzeptionelle Entwicklung einer Forschungsinfrastruktur für die e-Humanities in Deutschland“ am 22. Januar 2009 in Göttingen. In: LIBREAS. Library Ideas, 14 – URL: http://www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe14/016jah.htm
- JISC (2003) An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications (JISC Circular, 6 (03) (Revised)) – URL: <http://www.dcc.ac.uk/docs/6-03Circular.pdf>
- Kahn, Robert; Wilensky, Robert (2006) A framework for distributed digital object services. In: International Journal on Digital Libraries, 6 (2), S. 115-123 – URL: http://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf

- Klump, Jens (2009) Digitale Forschungsdaten. In: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine et al. [Hrsg.] (2009) nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.0, Kapitel 17.10, S. 104-115 – URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_292.pdf
- Kluttig, Thekla (2008) Bericht über das DFG-Rundgespräch „Forschungsprimärdaten“ am 17.01.2008 – URL: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprimaerdaten_0108.pdf
- Koltzenburg, Claudia (2008) Check-Listing digital objects in context. In: OCLC Systems & Services, 4 (24), S. 227-239 – URL: <http://82.109.207.14/Insight/ViewContentServlet;jsessionid=5F42C26D80596AD37284A78D2929AEE6?contentType=Article&Filename=Published/EmeraldFullTextArticle/Articles/1640240406.html>
- Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar [Hrsg.] (2004) Grundlagen der praktischen Information und Dokumentation (5. Auflage), K.G. Saur: München
- Kuhlen, Rainer (2004) Information. In: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar [Hrsg.] (2004) Grundlagen der praktischen Information und Dokumentation (5. Auflage), K.G. Saur: München, S. 3-20
- Ludwig, Jens (2008) Digitale Objekte und ihre Eigenschaften. In: o.V. (2008) nestor / DPE Summer School 2008. Einführung, Technologien und Strategien für die digitale Langzeitarchivierung. BDB-Musikakademie in Staufen/Breisgau vom 16. bis 20. Juni 2008
- Lynch, Clifford (2007) The shape of the scientific article in the developing cyberinfrastructure. CTWatch Quarterly, 3 (3) – URL: <http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/>
- Lynch, Clifford; Jacobs, Neil; Lagoze, Carl; Van de Sompel, Lagoze, (2007) The OAI-ORE effort: progress, challenges, synergies. In: International Conference on Digital Libraries archive. Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada
- Lynch, C. A. (2003) Institutional repositories. Essential infrastructure for scholarship in the digital age. (ARL Bimonthly Report, 226, S. 1-7) – URL: <http://www.arl.org/bm~doc/br226ir.pdf>
- Lyon, Liz (2007) Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report – URL: <http://www.jisc.ac.uk/publications/documents/dealingwithdatareportfinal.aspx>
- Lyon, Liz; Heery, Rachel; Duke, Monica et al. (2004) eBank UK. Linking research data, scholarly communication and learning. – URL: <http://eprints.soton.ac.uk/8183/>

- Margulies, Simon B. (2009) Digitale Daten als Quelle der Geschichtswissenschaft. Eine Einführung. Hamburg: Verlag Dr. Kovač
- Maron, Nancy L.; Smith, Kirby K. (2008) Current Models of Digital Scholarly Communication. Results of an Investigation Conducted by Ithaka Strategic Services for the Association of Research Libraries – URL: <http://www.arl.org/bm~doc/current-models-report.pdf>
- McCarty, Willard (2008) What's going on? In: Literary and Linguistic Computing, 3 (23) – URL: <http://llc.oxfordjournals.org/cgi/reprint/23/3/253>
- McCarty, Willard (2005) Humanities Computing. Basingstoke: Palgrave Mcmillan
- Meffert, Katja (2009) State-of-the-Art-Analyse. Studie über existierende (nationale, internationale) e-Humanities-Zentren, andere Initiativen bzw. Ansätze in diesem Bereich (Präsentation im Rahmen des e-Humanities-Abschluss-Workshops am 22.01.2009 in der SUB Göttingen) – URL: http://www.textgrid.de/fileadmin/TextGrid/konferenzen_vortraege/eHumanities_0109/e-humanities%20-%2002%20-%20State-of-the-Art%20Analyse%20-%20Meffert.pdf
- Meister, Jan Christoph (2005) Projekt Computerphilologie. Über Geschichte, Verfahren und Theorie rechnergestützter Literaturwissenschaft. In: Segeberg, Harro; Winko, Simone [Hrsg.] (2005) Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm Fink Verlag, S. 315-337 – URL: <http://www1.uni-hamburg.de/DigiLit//meister/computerphilologie.html>
- Mittelstraß, Jürgen [Hrsg.] (2005) Enzyklopädie Philosophie und Wissenschaftstheorie. 2., neu bearbeitete und wesentlich ergänzte Auflage. Stuttgart et al.: J. B. Metzler (3)
- Nelson, Michael L.; Van de Sompel, Herbert (2006) IJDL special issue on complex digital objects. Guest editors' introduction. In: International Journal on Digital Libraries 6 (2), S. 113-114 – URL: <http://dx.doi.org/10.1007/s00799-005-0127-y>
- Nentwich, Michael (2003) Cyberscience. Research in the age of the Internet. Vienna: Austrian Acad. of Sciences Press
- Nentwich, Michael (1999) Cyberscience: Die Zukunft der Wissenschaft im Zeitalter der Informations- und Kommunikationstechnologien. MPIfG Working Paper 99/6 – URL: <http://www.mpi-fg-koeln.mpg.de/pu/workpap/wp99-6/wp99-6.html>
- Nestor [Hrsg.] (2009) Digitale Forschungsdaten bewahren und nutzen. Für die die Wissenschaft und für die Zukunft (nestor Arbeitsgruppe Grid /e-sience und Langzeitarchivierung) – URL: http://www.langzeitarchivierung.de/downloads/bericht/nestor_bericht_forschungsdaten.pdf

- Neuroth, Heike (2009) eResearch und die Rolle der wissenschaftlichen Bibliotheken (Vortrag im Berliner Bibliothekswissenschaftlichen Kolloquium am 14.7.2009) – URL: <http://www.ibi.hu-berlin.de/institut/veranstaltungen/bbk/bbk-material/bbk-2009-neuroth>
- Neuroth, Heike; Oßwald, Achim; Scheffel, Regine et al. [Hrsg.] (2009) nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.0 – URL: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_20.pdf
- Neuroth, Heike (2008) Research Data Curation in eHumanities. Präsentation im Rahmen des <philtag>-Workshops am 13./14. Oktober in Trier „Communicating eHumanities: Archives, Textcentres, Portals” – URL: <http://germazope.uni-trier.de/Projects/KoZe2/aktuelles/Neuroth>
- National Science Board (2005) Long-lived digital data collections. Enabling research and education in the 21st century (National Science Foundation / National Science Board) – URL: http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf#search=%22long%20lived%20digital
- National Science Foundation (2007) Cyberinfrastructure Vision for 21st Century Discovery (National Science Foundation / Cyberinfrastructure Council) – URL: <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>
- Powell, Andy; Nilsson, Mikael, Naeve, Ambjörn (2007) DCMI Abstract Model – URL: <http://dublincore.org/documents/abstract-model/#DCRDF>
- Ockenfeld, Marlies [Hrsg.] (2008) Verfügbarkeit von Informationen. 30. Online-Tagung der DGI, 60. Jahrestagung der DGI. Frankfurt am Main (Tagungen der Deutschen Gesellschaft für Informationswissenschaft und -praxis (DGI-Tagungen), 11)
- OCLC [Hrsg.] (2003) New Flows of Scholarly Materials. OCLC Environmental Scan – URL: <http://www.oclc.org/reports/escan/research/newflows.htm>
- OAI-ORE (2008) ORE Specifications and User Guides– URL: <http://www.openarchives.org/ore/1.0/toc>
- o. V. (2008) nestor / DPE Summer School 2008. Einführung, Technologien und Strategien für die digitale Langzeitarchivierung (Materialien einer Veranstaltung in der BDB-Musikakademie in Staufen/Breisgau vom 16. bis 20. Juni 2008)
- Peter, Christian [Hrsg.] (2002) Die Geisteswissenschaften heute und morgen. Herbsttagung der Schweizerischen Akademie der Geistes- und Sozialwissenschaften, Bern, 15. November 2001. Bern: Schweizerische Akademie der Geistes- und Sozialwiss.
- Pfeiffenberger, Hans (2009) Umgang mit Forschungsdaten. Die Aktivitäten der AG Forschungsdaten in der Allianz der Wissenschaftsorganisationen (Präsentation im Rahmen des Bibliothekartags 2009 am 05.06.2009 in Erfurt)

- Pfeiffenberger, Hans (2008) Allianz AG Forschungsdaten. Auf dem Weg zu einer Policy im Umgang mit Forschungsdaten (Präsentation im Rahmen des KoLa-Wiss-Workshops am 27.01.2009 in Göttingen) – URL: http://kolawiss.uni-goettingen.de/workshop/ws2/Forschungsdaten_Pfeiffenberger.pdf
- Pfeiffenberger, Hans (2007) Offener Zugang zu wissenschaftlichen Primärdaten. In: Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB), 4-5 (54), S. 207-210 – URL: http://zs.thulb.uni-jena.de/servlets/MCRFileNodeServlet/jportal_derivate_00062527/j07-h4-5-auf-10.pdf
- Pradt Lougee, Wendy (2007) Promoting Digital Scholarship: What do Scholars Do? What do Scholars Need? – URL: <http://www.clir.org/activities/digitalscholar/Nov07Lougee.pdf>
- Rapp, Andrea (2007) Das Projekt „TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung - ein Community-Grid für die Geisteswissenschaften“. Chancen und Perspektiven für eine neue Wissenschaftskultur in den Geisteswissenschaften. In: Arbeitsgemeinschaft historischer Forschungseinrichtungen in der Bundesrepublik Deutschland [Hrsg.] (2007) Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland: Berichtsjahr 2006. München: Oldenbourg, S. 61-68 – URL: http://www.ahfmuenchen.de/Forschungsberichte/Jahrbuch2006/AHF_Jb2006_FB_B1_Rapp.pdf
- Research Data Strategy Working Group (2008) Stewardship of Research Data in Canada: A Gap Analysis (Research Data Canada) – URL: <http://data-donnees.gc.ca/docs/GapAnalysis.pdf>
- Research Information Network [Hrsg.] (2008) Discovering physical objects. Meeting researcher's needs. A Research Information Network Report, October 2008 – URL: <http://www.rin.ac.uk/objects>
- Rugemer, Christine (2008) Kinetik der Kulturen. In: Research*eu, 56, S. 29ff. – URL: <http://www.schattenblick.de/infopool/europool/wissen/euwif338.html>
- Sandkühler, Hans Jörg; Regenbogen, Arnim; Friemert, Chup (1990) Europäische Enzyklopädie zu Philosophie und Wissenschaften. Hamburg: Meiner
- Schirmbacher, Peter (2009) Virtuelle Forschungsumgebungen. Neue Formen der wissenschaftlichen Kooperation. Vortrag im Rahmen des DINI/DFG-Workshops „Virtuelle Forschungsumgebungen“ im Februar 2009 – URL: <http://www.dini.de/fileadmin/workshops/dfg-dini-forschungsumgebungen-2009/schirmbacher.pdf>
- Schirmbacher, Peter (2005) Die neue Kultur des elektronischen Publizierens. In: cms-journal, 27, S. 19-22 – URL: <http://edoc.hu-berlin.de/cmsj/27/schirmbacher-peter-19/PDF/schirmbacher.pdf>

- Schütt, Hans-Peter (2005) Der „Geist“ in den Geisteswissenschaften. In: Arnswald, Ulrich; Nida-Rümelin, Julian; Arnswald, Wilfried [Hrsg.] (2005): Die Zukunft der Geisteswissenschaften. [Wilfried Arnswald zum 70. Geburtstag]. Heidelberg: Manutius-Verl. (Schriften des European Institute for International Affairs, Heidelberg), S. 63-76
- Schreibman, Susan; Siemens, Ray; Unsworth John (2004) A companion to digital humanities. Malden, Mass.: Blackwell Publ. (Blackwell companions to literature and culture, 26) – URL: <http://www.digitalhumanities.org/companion/>
- Schweibenz, Werner (2008) Vom traditionellen zum virtuellen Museum. Die Erweiterung des Museums in den digitalen Raum des Internets. Frankfurt am Main: DGI (DGI-Schrift Informationswissenschaft, 11)
- Schoepflin, Urs (2009) Open-Access-Forschungsinfrastruktur für Kulturerbe [Vortrag im Berliner Bibliothekswissenschaftlichen Kolloquium am 14.7.2009] - URL: <http://www.ibi.hu-berlin.de/institut/veranstaltungen/bbk/bbk-material/bbk-2009-schoepflin>
- Segal, Jane; Spiro, Lisa; Francis, Pamela (2007) Impact of Digital Resources on Humanities Scholarship. – URL: <http://library.rice.edu/services/dmc/projects/the-impact-of-digital-resources-on-humanities-research>
- Segeberg, Harro; Winko, Simone [Hrsg.] (2005) Digitalität und Literalität. Zur Zukunft der Literatur. München: Fink - URL: <http://www1.uni-hamburg.de/DigiLit//index.html>
- Seiffert, Helmut (1997) Wörterbuch der wissenschaftstheoretischen Terminologie. Orig.-Ausg. München: Beck (Beck'sche Reihe, 1200)
- Sheehan, Mark C. (2008) Cyberinfrastructure. Changing a Cottage Industry. In: Educause Review 4 (43) – URL: <http://connect.educause.edu/Library/EDUCAUSE+Review/CyberinfrastructureChange/46970?time=1237890082>
- Snow, Kellie; Ballaux, Bart; Christensen-Dalsgaard, Birte et al. (2008) Considering the User Perspective. Research into Usage and Communication of Digital Information. In D-Lib Magazine, 5/6 (14) – URL: <http://www.dlib.org/dlib/may08/ross/05ross.html>
- Smith, Abby (2007) Thoughts on Scale and Complexity (NSF / JISC workshop on data-driven science, April 17-19, 2007) – URL: <http://www.sis.pitt.edu/~repwkschop/smith%20panel.pdf>
- Smith, Abby (2004) Preservation. In: Schreibman, Susan; Siemens, Ray; Unsworth John [Hrsg.] (2004) A companion to digital humanities. Malden, Mass.: Blackwell Publ. (Blackwell companions to literature and culture, 26), S. 576-591

- Sturm, Beate (2009) Publikation und Austausch von Forschungsdaten. Ergebnisse aus Umfragen des EU-Projekts PARSE.Insight (Präsentation im Rahmen des Bibliothekartags in Erfurt 2009 am 05.06.2009)
- Struve, Karen (2005) Von der Wirklichkeit zur Wissenschaft. Zum Verhältnis von Gegenstand und Methode in den Sprach-, Literatur- und Kulturwissenschaften. In: Sollte-Gresser, Christiane; Struve, Karen; Ueckmann, Natascha [Hrsg.] (2005) Von der Wirklichkeit zur Wissenschaft, Münster: LIT Verlag, S. 9-28
- Sukovic, Suzana (2008) Convergent Flows: Humanities Scholars and their Interactions with Electronic Texts. In: Library Quarterly, 3 (78), S. 263-284.
- Thibodeau, Kenneth (2002) Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years (Council on Library and Information Resources (CLIR)) – URL: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>
- Treloar, Andrew; Wilkinson, Ross (2008) Access to Data for eResearch: Designing the Australian National Data Service Discovery Services. In: The International Journal of Data Curation, 2 (3), S. 151-158 – URL: <http://www.ijdc.net/ijdc/article/view/95/113>
- UK Data Archive (2008) Managing and Sharing Data. A Best Practice Guide for Researchers – URL: <http://www.data-archive.ac.uk/news/publications/managingsharing.pdf>
- Uhlir, Paul F.; Schröder, Peter (2007) Open Data for a Global Science. In: Data Science Journal, 6, S. OD37-OD53
- Unsworth, John (2000) Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this? Part of a symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London, May 13, 2000 – URL: <http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html>
- Van de Sompel, Herbert (2008) An introduction to the ORE interoperability framework (4th Search&Find Workshop 22. August 2008, Gent, Belgien) – URL: http://public.lanl.gov/herbertv/presentations/ORE_Gent_hvds.pdf
- Van de Sompel, Herbert (2008) The OAI-ORE Interoperability Framework in the current Scholarly Communication Context – URL: <http://www.slideshare.net/hvdsomp/the-oaiore-interoperability-framework>
- Van de Sompel, Herbert; Lagoze, Carl, Nelson, Michael L. (2007) Open Archives Initiative Object Re-Use & Exchange – URL: <http://elag2007.upf.edu/papers/sompel.pdf>

- Van de Sompel, Herbert; Lagoze, Carl (2007) Interoperability for the Discovery, Use, and Re-Use of Units of Scholarly Communication. In: CTWatch Quarterly, 3 (3) – URL: <http://www.ctwatch.org/quarterly/articles/2007/08/interoperability-for-the-discovery-use-and-re-use-of-units-of-scholarly-communication/>
- Van Horik, René (2007) Data Curation. In: A DRIVER'S Guide to European Repositories. Inventory study of important DR related issues and good practices (D 7.2) – URL: http://www.driver-repository.eu/PublicDocs/D7.2_1.1.pdf
- Verhaar, Peter (2009) Enhanced Publications. Object Models and Functionalities (Driver Documentation) – URL: http://www.driverrepository.eu/component/option,com_jdownloads/Itemid,83/task,finish/cid,54/catid,8/
- Voss, Alexander & Procter, Rob (2009) Virtual research environments in scholarly work and communications. In: Library High Tech, 2 (27), S. 174-190 – URL: <http://www.emeraldinsight.com/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2380270202.pdf>
- Voß, Jakob (2009) Zur Neubestimmung des Dokuments im rein Digitalen. Erscheint in: LIBREAS. Library Ideas, 15 (in Vorbereitung)
- Warner, Simeon, Bekaert, Jeroen, Lagoze, Carl, et al. (2007) Pathways. Augmenting interoperability across scholarly repositories. In: International Journal on Digital Libraries (JODL), 1/2 (7), S. 35-52 – URL: <http://dx.doi.org/10.1007/s00799-007-0016-7>
- Warwick, Claire (2004) Print Scholarship and Digital Resources. In: Schreibman, Susan; Siemens, Ray; Unsworth John [Hrsg.] (2004) A companion to digital humanities. Malden, Mass.: Blackwell Publ. (Blackwell companions to literature and culture, 26), S. 366-382 – URL: <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-6&toc.depth=1&toc.id=ss1-4-6&brand=default>
- Warwick, Claire; Galina, Isabel; Rimmer, Jon et al. (2009) Documentation and the users of digital resources in the humanities. In: Journal of Documentation, 1 (56), S. 33-57
- Weenink, Kasja; Waaijers, Leo (2007) A DRIVER'S Guide to European Repositories. Five studies of important Digital Repository related issues and good Practices. Amsterdam. Amsterdam University Press – URL: <http://dare.uva.nl/document/93898>
- Weichselgartner, Erich (2008) Fünf Jahre Primärdatenarchivierung in der Psychologie. Ein Erfahrungsbericht. In: Ockenfeld, Marlies [Hrsg.] Verfügbarkeit von Informationen. 30. Online-Tagung der DGI, 60. Jahrestagung der DGI. Frankfurt am Main (Tagungen der Deutschen Gesellschaft für Informationswissenschaft und -praxis (DGI-Tagungen), 11), S. 259-267 – URL: <http://www.weichselgartner.de/edu/downloads/dgi2008.pdf>

- Winko, Simone (2005) Hyper – Text – Literatur. Digitale Literatur als Herausforderung an die Literaturwissenschaft. In: Segeberg, Harro; Winko, Simone [Hrsg.] (2005) Digitalität und Literalität. Zur Zukunft der Literatur. München: Wilhelm Fink Verlag, S. 137-157 – URL: http://www1.uni-hamburg.de/DigiLit/winko/hyper_text_literaturwissenschaft.html
- Wissenschaftsrat [Hrsg.] (2006) Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland. Wissenschaftsrat – URL: <http://www.wissenschaftsrat.de/texte/7068-06.pdf>
- Woutersen-Windhower, Saskia & Brandsma, Renze (2009) Enhanced Publications. State-of-the-art (Driver Documentation) – URL: http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,83/task,finish/cid,53/catid,8/
- Yakel, Elizabeth (2007) Digital Curation. In: OCLC Systems & Services: International digital library perspectives, 4 (23), S. 335-340
- Zimmermann, Harald H. (2004) Information in der Sprachwissenschaft. In: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar [Hrsg.] (2004) Grundlagen der praktischen Information und Dokumentation (5. Auflage), K.G. Saur: München, S. 705-709
- Zorich, Diane (2008) A Survey of Digital Humanities Centers in the United States. Prepared for the Council on Library and Information Resources (CLIR). Princeton. – URL: http://www.uvasci.org/wp-content/uploads/2008/06/dhc-survey-final-rept-2008_05_22.doc

10.2 Weitere Internetquellen

Alle Internetquellen wurden zuletzt am 02.08.2009 aufgerufen.

Atom – URL: <http://www.atompub.org/>

Alliance of Digital Humanities Organizations – URL: <http://digitalhumanities.org/>

Arts and Humanities Data Service (AHDS) – URL: <http://ahds.ac.uk/>

Bing Maps – URL: <http://www.bing.com/maps/>

CIDOC Conceptual Reference Model (CIDOC CRM) – URL:
<http://cidoc.ics.forth.gr/index.html>

Das deutsche Portal zum 7. EU-Forschungsrahmenprogramm – URL:
<http://www.forschungsrahmenprogramm.de/>

D-Grid-Initiative – URL: <http://www.d-grid.de/>

Digital Curation Center (DCC) – URL: <http://www.dcc.ac.uk/>

Discovery Project – URL: <http://www.discovery-project.eu/home.html>

DCMI Abstract Model – URL: <http://dublincore.org/documents/abstract-model/>

Digital Item Declaration Language – URL: <http://xml.coverpages.org/mpeg21-didl.html>

Digital Repository Infrastructure Vision for European Research (Driver) – URL:
<http://www.driver-repository.eu/> Driver Enhanced Publications – URL:
<http://www.driver-repository.eu/Enhanced-Publications.html>

Dublin Core (DC) – URL: <http://dublincore.org/>

European Strategy Forum for Research Infrastructures (ESFRI) – URL:
<http://cordis.europa.eu/esfri/>

Europeana – URL: <http://www.europeana.eu/portal/>

Entity Relationship Model – URL: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/daten-wissen/Datenmanagement/Daten-/Entity-Relationship-Model-->

FRBR – URL: <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>

Förderung der wissenschaftlichen Informationslandschaft in Deutschland – Der Aufbau Virtueller Forschungsumgebungen – URL:
<http://www.dini.de/veranstaltungen/workshops/forschungsumgebungen-2009>

Göttinger e-Humanities-Abschluss-Workshop (DFG) – URL:
<http://www.textgrid.de/konferenzen/e-humanities-abschluss-workshop-dfg.html>

Google Maps – URL: <http://maps.google.de/>

Hypernietzsche – URL: <http://www.hypernietzsche.org/>

Linked Data – URL: <http://www.w3.org/DesignIssues/LinkedData.html>

Metadata Encoding and Transmission Standard (METS) – URL:
<http://www.loc.gov/standards/mets/>

Open Archives Initiative Object Reuse and Exchange (OAI-ORE) – URL:
<http://www.openarchives.org/ore/>

Web Ontology Language (OWL) – URL: <http://www.w3.org/TR/owl-features/>

Resource Description Framework (RDF) – URL: <http://www.w3.org/RDF/>

Resource Description Framework in attributes (RDFa) –URL: <http://rdfa.info/>

Text Encoding Inscription (TEI) – URL: <http://www.tei-c.org/>

Text-Grid – URL: <http://www.textgrid.de/>

Unified Modeling Language (UML) – URL: <http://www.uml.org/>

World Wide Web Consortium (W3C) – URL: <http://www.w3.org/>

11 Anhang

Tabelle 1 Übersicht zum Wissenschaftsbetrieb in Hinblick auf e-Research und ggf. Spezifika der digitalen Geisteswissenschaften in dieser Masterarbeit in Anlehnung an Nentwich (1999) Cyberscience, Tabelle 1 „Wandel des Wissenschaftsbetriebs auf dem Weg zur Cyberscience“ – Vgl. <http://www.mpifg.de/pu/workpap/wp99-6/tabelle1.htm>

Wissenschaftliche Wertschöpfung		e-Research und ggf. Spezifika der digitalen Geisteswissenschaften in dieser Arbeit	
Wissenschaftliche Wertschöpfung	Organisation	Institutionalisierung	Vernetzte Einrichtungen, Data Center
		Technische Ausstattung	ressourcenbasierte e-Infrastructure (insb. Grid-Computing bzw. Data Grid) für: Resource Sharing: Datennetze, Tools, Speicher, Rechenleistung
		Projektakquisition, Projektmanagement	Nutzung kollaborativer Plattformen (Groupware, Web 2.0-Tools) virtuelle und interdisziplinäre Kooperation und Kollaboration auf Basis vernetzter Strukturen → Virtuelle Forschungsumgebungen
		Informationsbeschaffung	Digitale Bibliotheken, Datenzentren, Repositorien
	Wissensproduktion	Datengewinnung	Simulationen, Modellierung, virtuelle Realität, Experimente, Erhebungen, Digitalisierung kultureller Artefakte, Konversion, Verknüpfung von Ressourcen
		Datenverwaltung	Digital Curation (Data Curation & Digital Preservation) unter Beachtung des Data Life Cycle mit dem Ziel der Nutzbarkeit
		Informationsverarbeitung und Analyse	Entwicklung von Methoden der Formalisierung von Bedeutungszuweisungen (Semantic Web)
		Erkenntnisproduktion, Bedeutungsfindung	Sprachaussagen, Anwendung von Instrumenten und Methoden auf Gegenstände, reflexiver und diskursiver Vorgang Strukturelle und semantische Verknüpfung von Ressourcen
		Wissensrepräsentation	<i>Semantische und strukturelle Beschreibung digitaler Objekte; Modellierung komplexer digitaler Objekte</i>
		Kommunikation und Prozessierung von Wissen	Kooperation
Diskurs	Online-Tools, kollaborative Plattformen, Virtuelle Forschungsumgebungen, E-Mail, Peer Review, Verknüpfung von Ressourcen		
Publikation	Print und elektronische Veröffentlichungen		
Lehre/Ausbildung	Multimedia-Lehrmaterial, eLearning, virtuelle Universität		

Tabelle 2: Aggregationsstufen und Bezeichnungen aus verschiedenen Perspektiven

Gemeinsame Eigenschaften in Anlehnung an die Aggregationsstufe (Beispiele)	Begriffsumfang des „Dokuments“ aus dokumentarischer Sicht ¹⁸⁷	Bezeichnung nach einer Typologie von Repositorien (Perspektive: content- types) ¹⁸⁸	Sammelrichtlinie „Research Data and Materials“ der Bibliothek der Universität Newcastle ¹⁸⁹	Bezeichnung nach Erich Weichselgartner (Psychologie) ¹⁹⁰	Bezeichnung nach Urs Schoepflin ¹⁹¹	Bezeichnung nach Nentwich (1999) aus Sicht der „Wissensproduktion“
unbehandelte, originale Daten (Messergebnisse, Mitschnitte, Bilder)	Primärdokument (Quelle oder Originalquelle)	Rohdaten (raw data)	„primary materials“; Fakten, Beobachtungen, computergenerierte Ergebnisse, Messungen und Erfahrungen	Rohdaten (Messergebnisse)	Rohdaten (nicht bearbeitete Ressourcen, Originale)	Primärquelle / Primärdaten (Gesetze, statistische Daten, Experimentaldaten)
ausgewertete Daten (Tabellen, Grafiken, Zahlenmatrizen)	Sekundärdokument als Ergebnis eines Dokumentationsprozesse s (z. B. Referenzdatenbank oder Bibliographie)	Abgeleitete Daten (derived data)	Forschungsdaten (research data): Numerisch, deskriptiv oder visuell aufbereitet	Primärdaten (Zahlenmatrizen)	Editionen, Annotationen	k. A.
fixiertes Forschungsergebnis	Tertiärdokument als Ergebnis eines Dokumentationsprozesse s bezogen auf Sekundärdokumente	Preprint, Final drafts etc.	k. A.	Publikation	Publikation	Sekundärliteratur (Fachliteratur im Volltext)

¹⁸⁷ Vgl. Glossar zu Kühlen et. al [Hrsg.] (2004) Grundlagen der praktischen Information und Dokumentation, S. 28

¹⁸⁸ Vgl. Heery, Rachel & Anderson, Sheila (2004) Digital Repositories Review – URL: http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf, S. 13

¹⁸⁹ Vgl. The University of Newcastle Australia: Research Data and Materials Management Procedure – URL: <http://www.newcastle.edu.au/policy/000870.html>; u. a zitiert bei Neuroth (2008) Research Data Curation in eHumanities. Die Sammelrichtlinie einer australischen Bibliothek wurde ausgewählt, weil sich australische wissenschaftliche und infrastrukturelle Einrichtungen bereits auf vergleichsweise hohem Niveau dem Thema Primärdaten widmen (vgl. Pfeiffenberger (2009) Umgang mit Forschungsdaten)

¹⁹⁰ Vgl. Weichselgartner (2008) Fünf Jahre Primärdatenarchivierung in der Psychologie

¹⁹¹ Vgl. Schoepflin (2009) Open-Access-Forschungsinfrastruktur für Kulturerbe

11.1 Abbildungsverzeichnis

Abbildung 1: Lebenszyklus von Daten nach dem Modell des britischen Digital Curation Center ...	46
Abbildung 2: Obtaining Information from Data (OAIS-Primer Seite 2-4, Abb. 2-2).....	52
Abbildung 3: Das digitale Informationsobjekt im OAIS Reference Model.....	64
Abbildung 4: Das Konzept des Information Packages im OAIS Referenz-Modell.....	65
Abbildung 5: Beispiele von PDI-Typen im OAIS Referenz-Modell	67
Abbildung 6: ORE Resource Map aus dem ORE Primer: http://www.openarchives.org/ore/1.0/primer-images/ADM_basics.png	92

11.2 Tabellenverzeichnis

Tabelle 1: „Wandel des Wissenschaftsbetriebs auf dem Weg zur Cyberscience“	116
Tabelle 2: Aggregationsstufen und Bezeichnungen aus verschiedenen Perspektiven.....	117